

The international journal of science / 1 June 2023

nature

ATOMIC POWER

X-rays capture elemental and chemical characteristics one atom at a time

Skeleton keys
Single-atom editing of molecular backbones eases drug discovery

Full disclosure
Transparency on pathogen origins key to avoiding pandemics

Learning experience
How religion can influence education outcomes in Africa

VISIT…

LANZAROTE Caliente.com

Nature.2023.06.03

[Sat, 03 Jun 2023]

- **This Week**
- **News in Focus**
- **Books & Arts**
- **Opinion**
- **Work**
- **Research**

This Week

- **Why ideas of ‘planetary boundaries’ must uphold environmental justice** [31 May 2023]
  • Earth-system boundaries define a safe operating space for humanity. Accounting for the planet’s most vulnerable people provides a stark warning of the work still to be done.'
- **Why is India dropping evolution and the periodic table from school science?** [30 May 2023]
  • India’s curriculum body needs to explain why it has removed foundational topics from school textbooks.
- **Why I’m leading Pacific Islands students in the fight on climate change** [30 May 2023]
  • Our voices were crucial in persuading UN states to back legal obligations to act on climate change. Here’s how we did it.

- EDITORIAL
- 31 May 2023

Why ideas of ‘planetary boundaries’ must uphold environmental justice

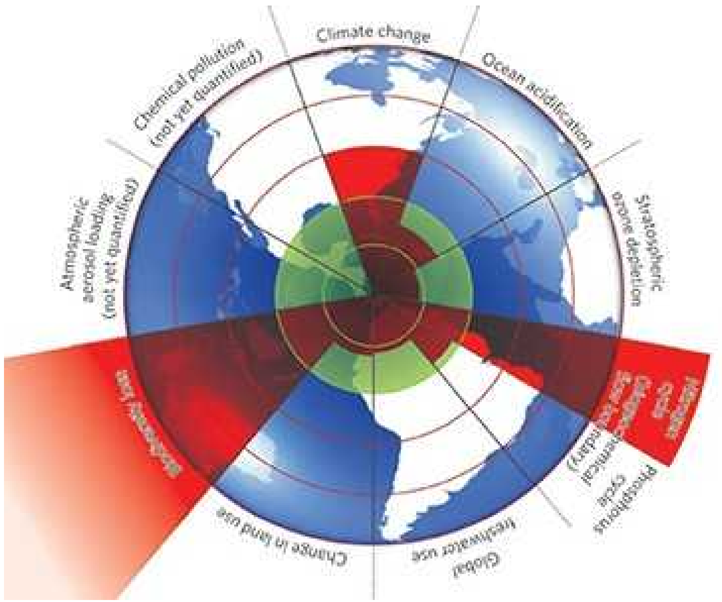
Earth-system boundaries define a safe operating space for humanity. Accounting for the planet’s most vulnerable people provides a stark warning of the work still to be done.'



Extreme weather, such as has occurred in Bangladesh, is having a disproportionate impact on vulnerable communities.Credit: Tanbir

Miraj/AFP/Getty

How many biophysical boundaries does our planet have? What are the limits of, say, carbon dioxide emissions, ocean acidification, chemicals and air pollution beyond which existence becomes unsafe for Earth and its inhabitants?



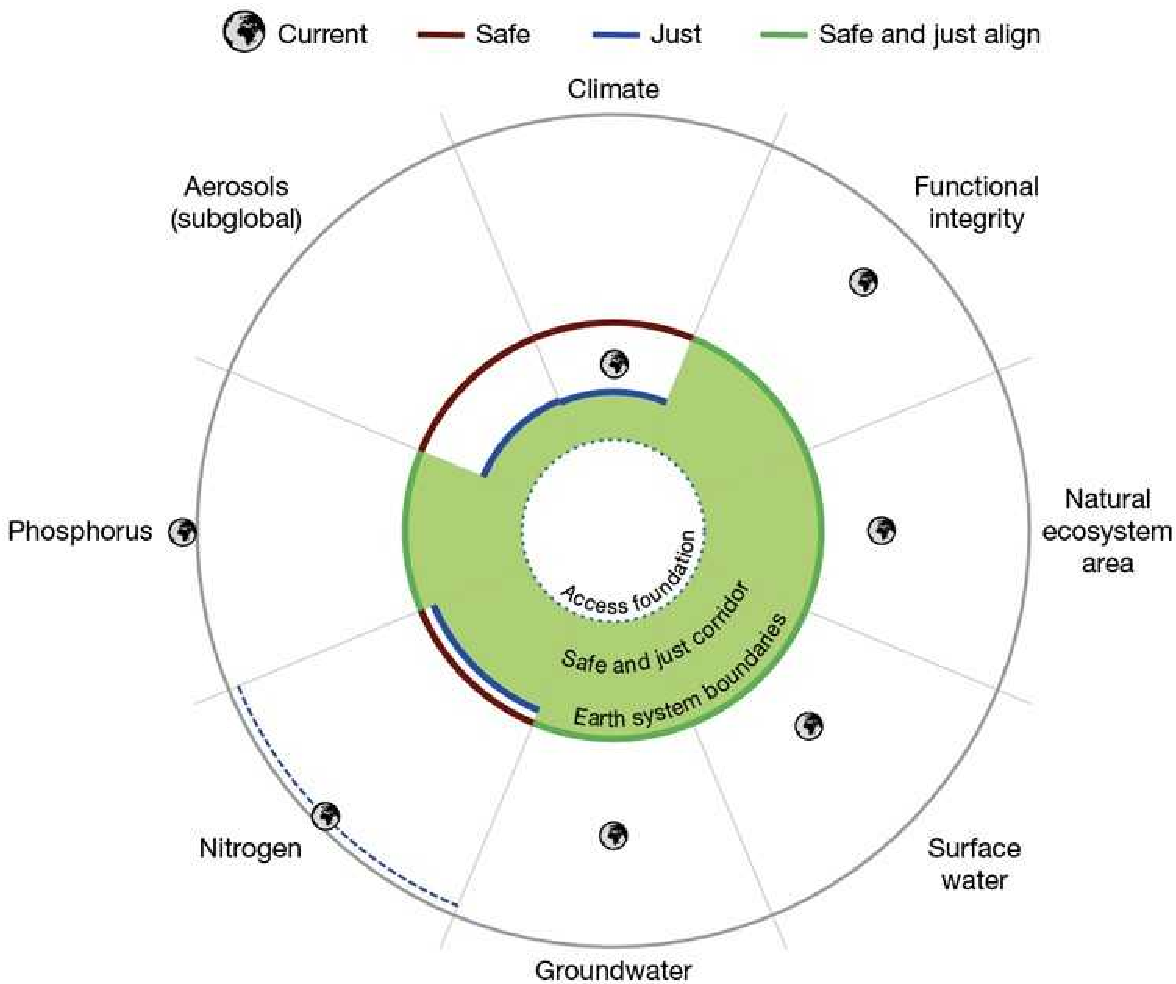
A safe operating space for humanity

Back in 2009, a team of researchers led by environmental scientist Johan Rockström grappled with these questions in an article published in *Nature* (J. Rockström *et al*. *Nature* **461**, 472–475; 2009). In the researchers' view, planet-altering human activities could be assembled into nine groups. Thresholds were calculated for most of them, beyond which the result could spell danger for the planet and its people. The scientists concluded that humanity has crossed three of these nine 'planetary boundaries', and that the remaining six would also be crossed unless remedial action was taken.

That article, called 'A safe operating space for humanity', has been extraordinarily influential in a relatively short time. Cities around the world have been experimenting with how to apply the findings, and researchers (including many in the original 2009 team) have continued to refine the planetary boundaries in response to feedback and new data.

A gap in the original concept was that it lacked environmental justice and equity — it needed to take into account the fact that everyone, especially the most vulnerable, has an absolute right to water, food, energy and health, alongside the right to a clean environment.

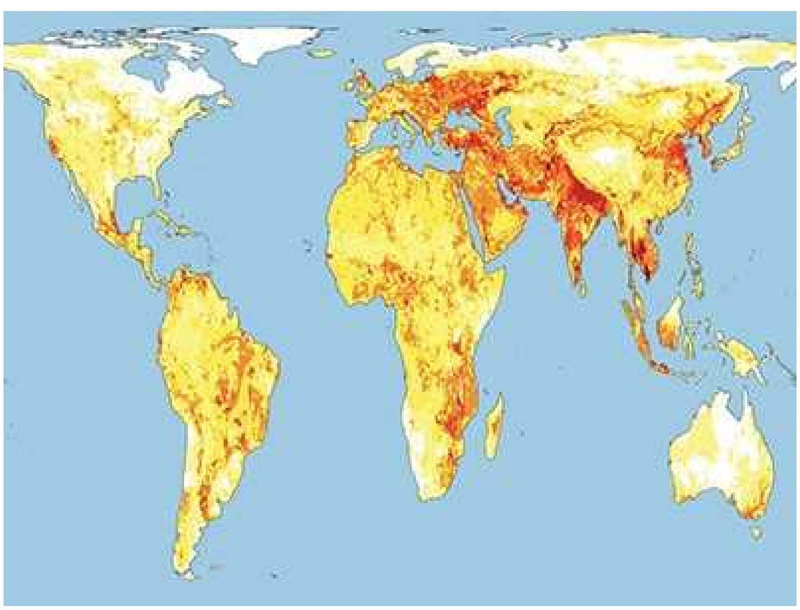
This week, Rockström, together with sustainability scientist Steven Lade and a team of researchers, have modified their original concept to incorporate justice alongside the biophysical boundaries. The resulting findings, which build on a study published in March in *Nature Sustainability* (J. Gupta *et al. Nature Sustain.* https://doi.org/grwfbk; 2023), show that seven out of eight thresholds have been crossed (see 'Planetary boundaries reboot'): the eight are climate, natural ecosystem area, ecosystem functional integrity, surface water, groundwater, nitrogen, phosphorus and aerosols.



Planetary boundaries reboot. The concept of ‘planetary boundaries’ has been updated to take into account the fact that everyone, particularly the most vulnerable, has an absolute right to water, food, energy and health, alongside the right to a clean environment. The red lines indicate a limit to what is ‘safe’ for the planet. The green space represents the threshold that is both safe for the planet and protects the world’s most vulnerable populations (‘safe and just’). The Earth-shaped icons show how, in seven of eight cases, thresholds for a safe and just world have already been crossed.Source: J. Rockström *et al*.

Stark warning

The findings are an even starker warning than were those reported in 2009. Arguably, the most striking change since 2009 is that the authors advocate that global warming should be limited to 1 °C above pre-industrial levels. This is lower than the 1.5 °C target agreed at the 2015 Paris climate conference. It is also a stricter constraint than the recommendation in the 2009 study to keep carbon emissions to 350 parts per million by volume (the pre-industrial value was 280 parts per million by volume).



Read the paper: Safe and just Earth system boundaries

The authors reason that keeping to 1.5 °C might well enable the world's more-affluent people to protect themselves, but it would create significant harm for the most vulnerable. The researchers estimate that some 200 million people would be exposed to unprecedented temperature increases and that 500 million people would be exposed to long-term sea-level rise.

In incorporating ideas of justice into their research, Rockström and colleagues build on a body of recent work. Not long after the 2009 paper was published, justice and equity were included in discussions that led to the United Nations Sustainable Development Goals (SDGs), announced in 2015. Environmental justice is also at the core of an innovative idea called doughnut economics. In 2012, economist Kate Raworth, then working for the aid agency Oxfam, and colleagues adapted the 2009 study's findings to include what Raworth called a "safe and just space". This space was represented by an area enclosed by a double circle, the 'doughnut', whose boundaries could not be crossed. Such a threshold would be measured using the indicators and methodology underpinning the SDGs.



How to identify unjust planetary change

The latest study shows how members of the 2009 team, working with a new generation of scientists and a more multinational team, have risen to the challenge of triangulating their original work with doughnut economics and the SDGs. This has not been easy, and the new work is very much an initial

step. In an accompanying News & Views article, Stephen Humphreys, who studies law and social justice at the London School of Economics, acknowledges the difficulty of setting numerical values when integrating ideas from the natural and social sciences. Readers can see how this process worked: alongside the paper, we are publishing the full discussion between authors and reviewers (J. Rockström *et al. Nature* https://doi.org/10.1038/s41586-023-06083-8; 2023). It underscores where there is agreement and disagreement, and where more data and further refinement will be needed.

This is a wonderful example of how well science functions when different teams study and refine each others' work. But there is a troubling aspect to the new findings, too. If seven of the eight thresholds have been crossed, what does that mean for our still-feeble efforts to move to a more sustainable path?

Researchers vary widely in their views on how this question should be addressed — from those who advocate working within the current economic system (known as green growth) to those arguing that the present economic system was itself a factor (if not the defining factor) in bringing about the present situation and requires transformation (known as post-growth or degrowth). Some months ago, we urged scientists representing these different approaches to forge more channels of communication between them.

The paper we are publishing represents one such opportunity. If the findings are anything to go by, there is no time to lose.

Nature **618**, 7 (2023)

doi: https://doi.org/10.1038/d41586-023-01749-9

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01749-9

- EDITORIAL
- 30 May 2023

Why is India dropping evolution and the periodic table from school science?

India’s curriculum body needs to explain why it has removed foundational topics from school textbooks.



India’s schools curriculum body has not consulted with parents, teachers and researchers on curriculum changes.Credit: Abbie Trayler-Smith/Panos Pictures

“To develop the scientific temper, humanism and the spirit of inquiry and reform.”

These words were crafted in 1976 in an amendment to India’s constitution. The constitution’s writers rightly saw the pursuit of evidence, reasoning and humanity as the responsibility of every citizen as India emerged from arguably the most tumultuous period in its history since gaining independence from Britain nearly three decades earlier.



India cuts periodic table and evolution from school textbooks — experts are baffled

But those attributes now seem to be less valued, at least by those involved in setting the country’s education policies. A series of changes to school science teaching have resulted in the deletion of the periodic table, explanations of evolution and electromagnetism, and discussions about the sustainable use of natural resources from the textbooks used by children aged 14–16.

These and other topics were removed from the curriculum last year to help lighten students’ workloads during the COVID-19 pandemic. But they have now been removed from textbooks, too. The National Council of Educational Research and Training (NCERT), the government-funded but operationally autonomous body tasked with producing India’s textbooks, has

not discussed the changes — which will affect more than 38 million children — with parents, teachers or researchers. Those who study science education have told *Nature* that they're baffled, not least by the lack of any engagement.

Pride in India

NCERT says that 'rationalization' is needed when content overlaps with material covered elsewhere in the curriculum, or when it considers content to be irrelevant. Moreover, India's 2020 National Education Policy says that students need to become problem-solvers and critical thinkers, and it therefore advocates less memorization of content and more active learning.



Indian scientists shocked as government scraps nearly 300 awards

NCERT also wants "a rootedness and pride in India, and its rich, diverse, ancient and modern culture and knowledge systems and traditions". Some people interpret this as a motivation to remove the likes of Charles Darwin and Michael Faraday, and instead use the time to learn more about India's precolonial history of science.

India is not the only postcolonial country grappling with the question of how to honour and recognize older or Indigenous forms of knowledge in its

school curricula. New Zealand is trialling the teaching of Māori ‘ways of knowing’ — *mātauranga Māori* — in a selection of schools across the country. But it is not removing important scientific content to accommodate the new material, and for good reason.

The process of evolution by natural selection and the principles underlying the periodic table are both fundamental concepts that explain — and encourage students to wonder about — the world at large. Life, in all its magnificent permutations and combinations, is the product of evolutionary processes. Meanwhile, a surprisingly small set of chemical elements form the building blocks of our physical world. How and why these two realms are the way they are can be traced back to lessons set out in conceptual frameworks that NCERT has axed.

Alternative methods

Moreover, there’s a great deal of literature and teaching practice that outlines how scientific concepts can be taught using visual methods, such as videos and animation, instead of by rote learning. An alternative is to embed concepts using non-fiction narrative storytelling. In the case of the periodic table, this would involve detailing how individual elements found their places in the table; the highs and lows, dramas and inflection points as researchers sought to get to the truth and be recognized for their achievements.



India pledges $4 billion for green energy to cut carbon emissions

Learning core scientific concepts, practising problem-solving and delving deep into the history of science — both local and global — needn't be done in isolation. The development of a scientific temperament and pride in heritage can go hand in hand. As we have written in these columns before, research does not advance without a firm grasp of what came before. In short, science and history complement each other.

Researchers who study India's education policy have told *Nature* that organizations that are critical of science are advocating for or influencing these changes to textbooks. They point to one organization in particular: the Rashtriya Swayamsevak Sangh, which has close ties to the ruling Bharatiya Janata Party.

NCERT does, of course, need to listen to opinions from the full community spectrum. But, as an autonomous body, it must be free to make its own decisions, and should always do so on the basis of the best available evidence. Public confidence in its decisions will be helped if it engages with all users: pupils, teachers, parents and researchers. Not doing so fuels all kinds of speculation, some of which might not be accurate.

NCERT needs to end its vow of silence. Few people would take issue with its ambition to boost critical thinking and promote learning by doing, or with

its desire for students to enjoy their education. Both can go hand-in-hand with exploring India’s rich pre- and postcolonial history of discovery and innovation.

Nature **618**, 8 (2023)

doi: https://doi.org/10.1038/d41586-023-01750-2

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01750-2

- WORLD VIEW
- 30 May 2023

Why I’m leading Pacific Islands students in the fight on climate change



Our voices were crucial in persuading UN states to back legal obligations to act on climate change. Here’s how we did it.

- Cynthia Houniuhi

On 29 March, I stayed up until 1 a.m., for good reason. From Sydney, Australia, where I am studying law, I watched live coverage of the United Nations General Assembly in New York City. More than 130 nations co-sponsored a resolution to put a case before the International Court of Justice (ICJ) on countries’ legal obligations to combat climate change. For the first time, the world’s highest court will rule on legal consequences for nations that are damaging the climate “by their acts and omissions”. The clout of

the ICJ's opinion could drive countries to strengthen their climate plans, and can be cited in domestic legal cases, although it is not legally binding.

As president of the advocacy group Pacific Islands Students Fighting Climate Change, I was ecstatic. The adoption of the resolution is a huge moment — the culmination of four years of hard work and dedication from us and our partners, including the government of Vanuatu and our umbrella organization, World's Youth for Climate Justice. This is a great step for climate justice and the rights of future generations. The adoption of the resolution surpassed our wildest dreams.

Our journey began in a classroom in Vanuatu at the University of the South Pacific School of Law and Social Science: 27 law students from eight Pacific Island countries banded together when our professor challenged us to research the most progressive legal pathways to address climate change and to propose one at the 2019 Pacific Island Leaders Forum Meeting in Tuvalu. We chose to agitate for an ICJ ruling because it was the most ambitious action on our list. And in the face of an existential threat to our people, ambition is what we need.



COP27: Rich countries must stop stalling talks on climate loss and damage compensation

Each of us brought to the table our stories of living in Pacific Island nations on the front line of global warming. I grew up in the Solomon Islands. Climate change is not a dinner-table conversation for my people — we do not have that luxury. It is what we lie awake worrying about at night. Climate change is the deaths of our people, whose losses we feel; it is lost possessions and homes, swept away by storms; it is wanting to move back to our childhood homes, but knowing they will be flooded in a matter of years; it is the threat of losing the places that our identities are tied to.

In the Pacific Islands, we contribute almost nothing to climate change, but we bear the brunt of it. Communities in Vanuatu lived through Cyclone Pam in 2015, which took more than two years to recover from, and during that time two category-4 tropical cyclones hit just days apart. Fiji dealt with one of the worst storms on record in 2016. In 2022, Tonga experienced the eruption of an underwater volcano. On the news, these are 'natural disasters' — but there is nothing natural about their frequency.

As a child, I was surrounded by nature, and my people relied on the environment to sustain us. I learnt how to fish, which plants are edible and our cultural dances that tell stories of people interacting harmoniously with the environment. I understood the value of the land and ocean to my people, and the value that people can bring to the land. Our traditional clothing comes from the land and the ocean: when I graduate, I plan to wear the rorochara — a headpiece made of seashells from the ocean that kisses the shorelines of our lands — that my grandmother wore when she got married.

It is heartbreaking that the journey towards solutions on the international level does not reflect the pace at which sea-level rise, intense rainfalls and storms are hitting my islands and my people.



When will global warming actually hit the landmark 1.5 ºC limit?

At least 5 of my country's 992 islands have been swallowed by the rising seas, and another 6 lost large swathes of their land, destroying entire villages. Some communities are losing sacred lands where their families are buried.

What finally pushed me into advocacy work was witnessing the loss of land to rising sea levels in Fanalei, a 2-kilometre wide island of the Solomon Islands. It was home to about 500 people, including some of my family. When I travelled back to the island during holidays, I loved seeing smiling and laughing kids playing football on the beach. It is a beautiful memory. Yet this image is fading as that island disappears before my eyes. When I visited in 2021, to bury my grandfather, food and water were scarce. At least 80% of the island's people have been forced to relocate. Only a few are left on the island.

One day, I will pass my grandmother's rorochara on to my future children. I hope they get to see the place where it came from, but that is not guaranteed. I imagine having nothing to show them but a photograph — and that daunting possibility drives me in this climate fight. I want to be able to look my children in the eyes when they ask, 'Did you do anything to help address this?' and say yes.

In this crucial decade, the world needs transformational change that accelerates climate action. We need guidance, and there is no better than that from the world's highest court. This campaign is my part, and it's an open opportunity for others to join in by lobbying their leaders for legal action on climate change or attending the public hearing that the ICJ is likely to hold in the next year. Look at your home and imagine that it is disappearing under water, as mine is. Ask yourself what action that motivates you to take.

Our world is an island, and it is drowning because of inaction.

Nature **618**, 9 (2023)

doi: https://doi.org/10.1038/d41586-023-01751-1

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01751-1

News in Focus

- **Gravitational-wave detector LIGO is back — and can now spot more colliding black holes than ever** [24 May 2023]
  • The twin gravitational-wave detectors have started a new observation run after a major upgrade.
- **Thousands protest Mexico's new science law** [19 May 2023]
  • Researchers are preparing to march against legislation that some say could harm basic science.
- **Do scientific meetings matter? Turning up for talks brings surprise benefits** [22 May 2023]
  • Talks that conference attendees could see in person are more likely to be cited than talks they most likely missed.
- **Does the roar of rocket launches harm wildlife? These scientists seek answers** [23 May 2023]
  • With rocket lift-offs set to increase drastically, a team will monitor the effects of noise pollution at a California spaceport.
- **Game-changing obesity drugs go mainstream: what scientists are learning** [23 May 2023]
  • Studies tackle who's most likely to lose weight on the new generation of anti-obesity medications.
- **Brain–spine interface allows paralysed man to walk using his thoughts** [24 May 2023]
  • The device provides a connection between the brain and spinal cord, allowing thought to control movement.
- **JWST spots biggest water plume yet spewing from a moon of Saturn** [18 May 2023]
  • The huge watery cloud spurting from Enceladus could carry the ingredients for life farther into space than previously known.
- **When will global warming actually hit the landmark 1.5 ºC limit?** [19 May 2023]

• The planet is on track to reach the 1.5 ºC average by the 2030s — although a new report suggests a single year will probably cross the line much sooner.

- **‘Almost magical’: chemists can now move single atoms in and out of a molecule’s core** [31 May 2023]
  • An explosion of skeletal editing methods to insert, delete or swap individual atoms in molecular backbones could accelerate drug discovery.

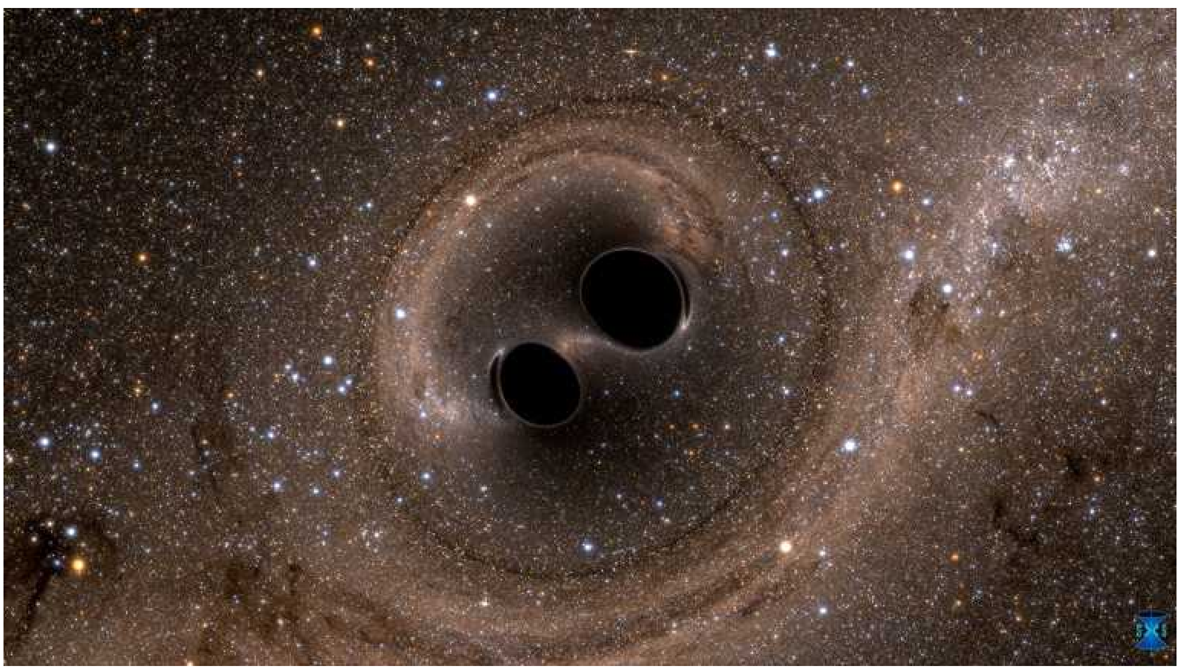
| Next section | Main menu | Previous section |

- NEWS
- 24 May 2023

Gravitational-wave detector LIGO is back — and can now spot more colliding black holes than ever

The twin gravitational-wave detectors have started a new observation run after a major upgrade.

- Davide Castelvecchi

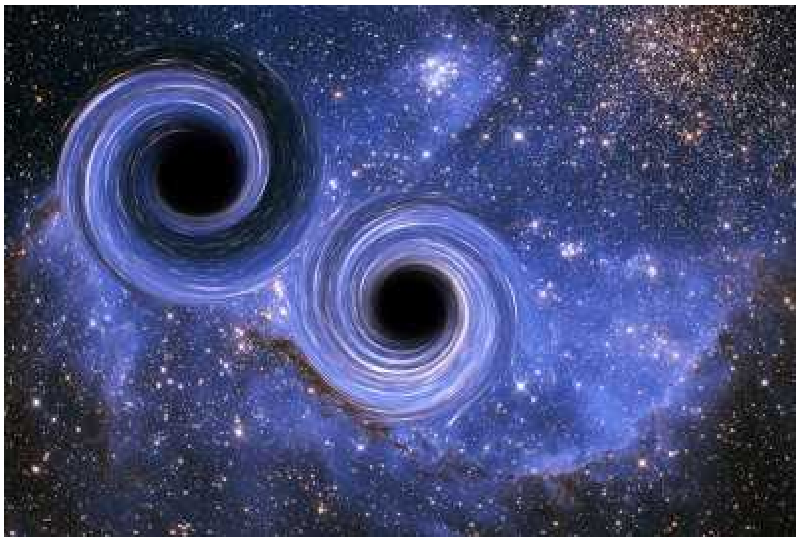


LIGO can detect gravitational waves that are generated when two black holes collide. Credit: The SXS Project

After a three-year hiatus made longer by pandemic troubles, the search for gravitational waves — ripples in space-time that are the hallmarks of

colliding black holes and other cosmic cataclysms — has resumed.

The Laser Interferometer Gravitational-Wave Observatory (LIGO), which has two massive detectors in Hanford, Washington, and Livingston, Louisiana, is now restarting with improved sensitivity after a multimillion-dollar upgrade. The improvements should allow the facility to pick up signals from colliding black holes every two to three days, compared with once a week or so during its previous run in 2019–20.



Gravitational-wave observatory LIGO set to double its detecting power

The Virgo detector near Pisa, Italy, which has undergone its own €8.4-million (US$9-million) upgrade, was meant to join in, but technical issues are forcing its team to extend its shutdown and perform further maintenance. “Our expectation is we’ll be able to restart by the end of summer or early autumn,” says Virgo spokesperson Gianluca Gemme, a physicist at Italy’s National Institute for Nuclear Physics in Genoa.

KAGRA, a gravitational-wave detector located under Mount Ikenoyama, Japan, is also restarting on 24 May. Its technology, although more advanced — it was inaugurated in 2020 — is being fine-tuned, and its sensitivity is still lower than LIGO’s was in 2015. Principal investigator Takaaki Kajita, a Nobel Prize-winning physicist at the University of Tokyo, says that KAGRA will join LIGO’s run for a month and then shut down again for another period of commissioning. At that point, the team will cool the

interferometer’s four main mirrors to 20 kelvin, Kajita says — a feature that sets KAGRA apart from the other detectors that will serve as the model for next-generation observatories.

Black-hole mergers

Gravitational waves are produced by large, accelerating masses, and the waves cyclically stretch and compress the fabric of space as they travel. Starting with LIGO’s historic first detection in 2015, most of the 90 or so gravitational-wave events recorded so far have been from the spiralling motion of pairs of black holes in the process of merging into one; a handful have been produced similarly by the merger of two neutron stars or a neutron star and a black hole.

LIGO, Virgo and KAGRA are all based on the same interferometer concept, which involves splitting a laser beam into two and bouncing the resulting beams between two mirrors at either end of a long vacuum pipe. (At LIGO, the two ‘arms’ of the interferometer are each 4 kilometres long; at Virgo and KAGRA, they are 3 km.) The two beams then come back and are made to overlap at a sensor in the middle. In the absence of any disturbances to space-time, the beams’ oscillations cancel one another out. But the passage of gravitational waves causes the arms to change in length with respect to each other, so that the waves don’t overlap perfectly, and the sensor detects a signal.



The LIGO detector in Livingston, Louisiana, is one of a pair based in the United States.Credit: Xinhua/Caltech/MIT/LIGO Lab

Typical gravitational-wave events change the length of the arms by only a fraction of the width of a proton. Sensing such minute changes requires painstaking isolation from noise coming from the environment and from the lasers themselves.

In upgrades carried out before the 2019–20 run, LIGO and Virgo tackled some of this noise with a technique called light squeezing. This approach deals with inherent noise caused by the fact that light is made of individual particles: when the beams arrive at the sensor, each individual photon can arrive slightly too early or too late, which means that the laser waves don't overlap and cancel out perfectly even in the absence of gravitational waves.

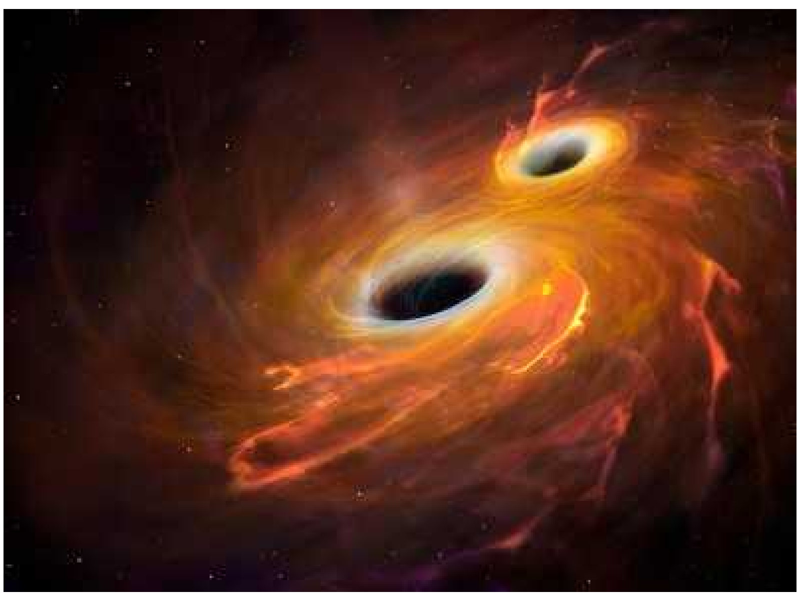
"It's like dropping a bucket of BBs [lead pellets]: it's going to make a loud hiss, but they all hit randomly," physicist Lee McCuller explained while showing a prototype of the LIGO interferometers at the Massachusetts Institute of Technology (MIT) in Cambridge. Light squeezing injects an auxiliary laser beam into the interferometer that reduces that effect. "Its

photons arrive more regularly, with less noise," said McCuller, who is now at the California Institute of Technology in Pasadena.

Quantum complications

The implementation of light squeezing has helped LIGO and Virgo to improve the detectors' sensitivities to higher-frequency gravitational waves.

But because of the bizarre rules of quantum mechanics, reducing the uncertainty in the arrival time of the photons increases random fluctuations in the laser waves' intensity. This causes the lasers to push on the interferometer mirrors and make them jitter, adding a different type of noise and potentially reducing their sensitivity to low-frequency gravitational waves. This is a "beautiful manifestation of nature", says MIT experimentalist Nergis Mavalvala, who helped to lead the development of the squeezing technology. "You cannot make an infinitely precise measurement: you have to pay the price somewhere else," she says.



What 50 gravitational-wave events reveal about the Universe

To deal with this issue, an important change in the most-recent upgrades of both LIGO and Virgo has been to build extra 300-metre-long vacuum pipes with mirrors at the ends, to store the auxiliary 'squeezing' beam for 2.5 milliseconds before injecting it into the interferometer. The role of these

pipes is to shift the waves of the auxiliary laser by distinct amounts depending on their wavelengths. This means that squeezing will be selective: it will decrease the noise at high frequency while also reducing mirror jitter at low frequencies.

MIT physicist Victoria Xu was part of the team that fine-tuned the new squeezing system at LIGO's Hanford laboratory, and she recalls the pleasant surprise when it was first turned on last November. "Things worked almost exactly as you might expect," she says.

Collapsing stars

With the improved sensitivity of the detectors, researchers will be able to extract more-detailed information about the spiralling objects that produce gravitational waves, including how each spins around its axis and how they revolve around each other. This means putting Albert Einstein's general theory of relativity — which predicts the existence of both black holes and gravitational waves — to stricter tests than ever before. The sheer number of observations will improve the big picture of how, and how often, black holes form from massive stars that collapse in on themselves.

Astrophysicists also anticipate that gravitational waves will reveal distinct types of signal in addition to those from black-hole mergers. One major hope is to pick up the gravitational signal of a collapsing star before it manifests as a supernova explosion — a feat that will be possible only if the collapse occurs somewhere in the Galaxy. Another ambition is to sense the continuous gravitational waves produced by ruggedness in the surface of a pulsar, a spinning neutron star that emits pulses of radiation.

The family of interferometers is due to expand by the end of the decade. The Indian government has announced that it will fund LIGO-India, a replica of the US observatories to be built in part with LIGO's spare components.

Nature **618**, 13-14 (2023)

doi: https://doi.org/10.1038/d41586-023-01732-4

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01732-4

- NEWS
- 19 May 2023

Thousands protest Mexico’s new science law

Researchers are preparing to march against legislation that some say could harm basic science.

- Myriam Vidal Valero



Opposition politicians hold a protest in Mexico’s Senate.Credit: Luis Barron/Eyepix Group/Future Publishing via Getty

More than 14,000 people have signed a letter protesting against the approval of a new science law in Mexico on 29 April. Researchers are organizing a march against it later this month and even calling for a strike. They say that the legislation — the General Law on Humanities, Sciences, Technologies and Innovation — consolidates power over science with the government and ignores the wishes of the research community. And they are angry that Mexico's ruling party, that of President Andrés Manuel López Obrador, rushed the law through — the approval didn't follow normal parliamentary procedures, the researchers say.



Frustration builds over lengthy delay in revamping Mexico's science law

The process was "atypical" and "irregular" says Fidel Sánchez, a bioinformatics researcher at the National Autonomous University of Mexico in Mexico City.

The law's passage adds to tensions between the research community and Obrador's administration, which has slashed science funding in Mexico and accused some scientists of organized crime.

The law has been four years in the making. Shortly after Obrador took office in 2018, the country amended its constitution to guarantee people the right to "enjoy the benefits" of science and technology. Officials then proposed updating Mexico's science law to align with this idea.

The scientific community worried that the country's main science agency, the National Council of Science and Technology (Conacyt), would dominate the process. So a series of discussion sessions were organized at which academic researchers and directors of educational institutions could express their views. Delays ensued because of the COVID-19 pandemic and disagreements among parties. This year, the government proposed holding seven official discussion forums, but only a few actually happened before last month's vote rushed the legislation through, alongside 19 other laws in a fast-tracked process.

Controversial changes

Although it sounds like a positive thing to guarantee the right to enjoy the benefits of science, Sánchez says a big problem with the legislation is that it is vague about how that will be ensured. Another issue is that, according to the law, a government council will prioritize science projects on the basis on their potential to solve national problems. This could diminish funding for basic science, Sánchez says. "How are we going to have a law that allows us to comply with the constitutional right that people have to education, science and its benefits, if we do not have the freedom of how we do science?"



Attempt to charge Mexican scientists with 'organized crime' prompts international outcry

That council will comprise the director of Conacyt — which the law transforms into the National Council for the Humanities, Science and Technology (Conahcyt) — as well as representatives of several government ministries and eight members of the scientific community. An internal advisory body, chaired by Conahcyt's director, will vet the scientists. This structure doesn't guarantee a "plurality of voices", says Alma Cristal, a chemist and president of the Mexican Society for the Advancement of Science in Mexico City.

The law also specifies that if a project receives funding from the government, any intellectual property that results will be owned by the government unless otherwise agreed, Cristal says. The problem with Conahcyt keeping the intellectual property "is that it doesn't have the mechanisms, facilities or the human or material resources to take that intellectual property and transfer it into products", she adds.

Researchers also worry that the new law introduces a change affecting science spending. The previous law specified that 1% of the country's gross domestic product should be invested in science, although Mexico never met that goal. Now the law specifies "that you can't invest less than the previous year", Sánchez says. This means that if the government increases the amount by only one peso (US$0.057) per year, it will be complying with the law. "There is no obligation to grow."

In a press release issued the day after the law went into effect, Conahcyt said that the legislation "is the result of an extensive 4-year process of dialogue and consensus, in which more than 70,000 people and institutions participated". The agency did not respond to *Nature*'s request for comment by the time this story was published.

High points

Not all scientists are against the new science law. Juan Martínez, a biologist at the Institute of Ecology in Xalapa, says some of his colleagues are "totally in favour".



[Mexican science suffers under debilitating budget cuts]

Others say the law has high points, even if they disagree with its overall direction. Cristal, for example, says its wording is more inclusive, and it mentions the aim of placing scientific staff in the country's embassies. It includes "some good wishes related to scientific diplomacy", she says, but doesn't specify how to achieve them.

Given the discontent with the law, it will probably be challenged in court. People and institutions have 30 days from 8 May, when it was enacted, to file suit. Martínez expects that the law will be modified: "If the judges consider that the constitution hasn't been respected, they are going to suggest to the legislature some kind of correction."

Nature **618**, 14-15 (2023)

doi: https://doi.org/10.1038/d41586-023-01665-y

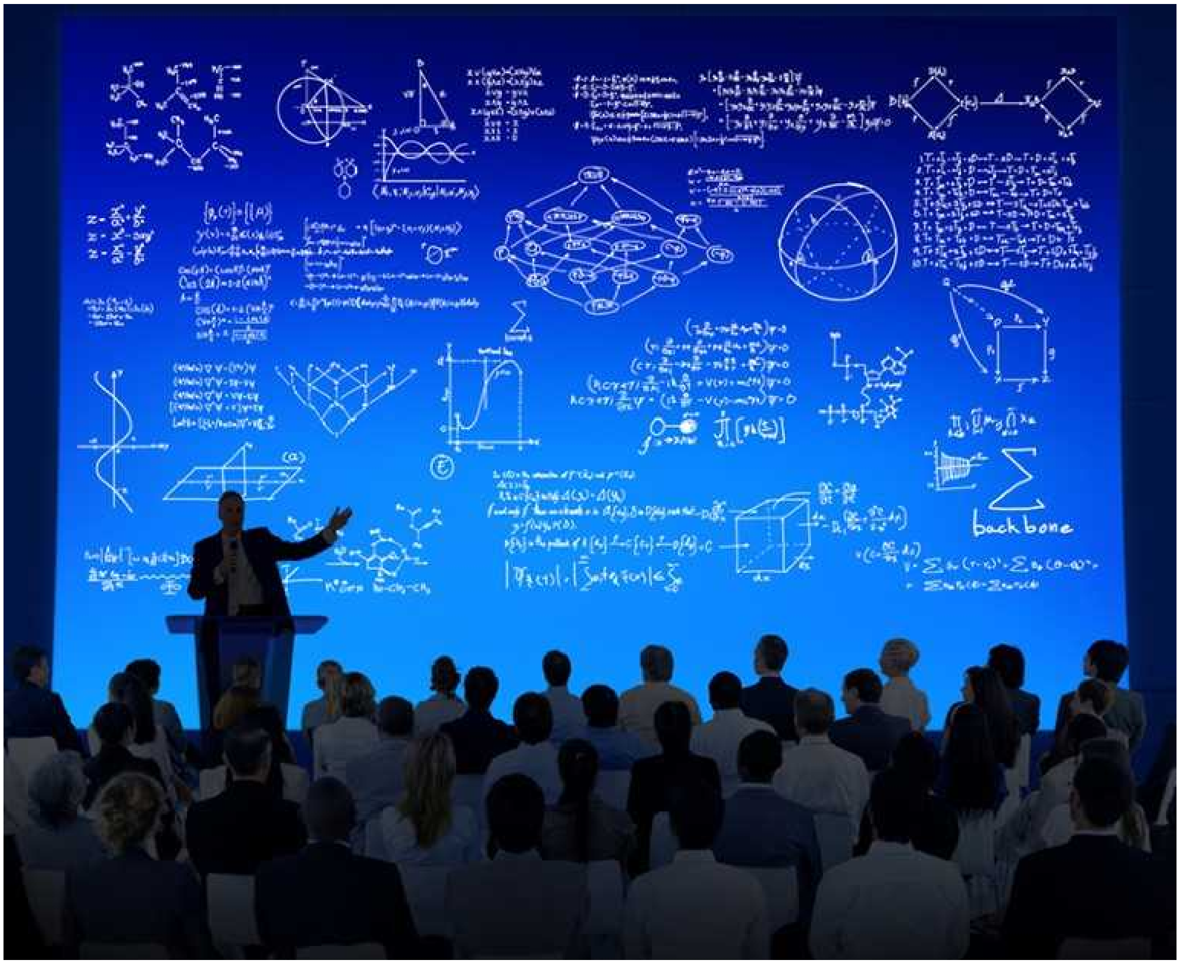
This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01665-y

- NEWS
- 22 May 2023

Do scientific meetings matter? Turning up for talks brings surprise benefits

Talks that conference attendees could see in person are more likely to be cited than talks they most likely missed.

- Katharine Sanderson



Seeing presentations in person raises the chance that a scientist will cite the work featured in the talk.Credit: Rawpixel/Shutterstock

Scientists who have attended meetings are more likely to cite work discussed in talks they could see in person, compared with results described in sessions that they could not attend[1]. That citation bump from in-person attendance accrues even for talks that conference attendees hadn't planned on listening to.

Attending a talk is "really, really effective" for increasing the chance that researchers will cite the work, says study co-author Misha Teplitskiy, an information scientist at the University of Michigan in Ann Arbor. The study was posted on the arXiv preprint server on 5 May. It has not yet been peer reviewed.

Teplitskiy has long been interested in where scientists get their ideas, but struggled to measure this. Then he was introduced to David Karger, a computer scientist at the Massachusetts Institute of Technology in Cambridge, who had developed an app called Confer to help computer scientists plan their personal schedules at conferences. Scientists using the app can 'like' a talk by clicking on it.

Showing up

Teplitskiy, Karger and their colleagues used data from the app to deduce which talks 2,404 Confer users attended at 25 computer science conferences held between 2013 and 2020. The team assumed that a conference participant was more likely to attend a 'liked' talk if it took place in a session that did not conflict with anything else on the attendee's schedule.

The authors then assessed the works cited by those conference attendees within two years of the meeting. They found that for attendees who had multiple schedule conflicts with a liked talk, the probability of later citing that talk was 2.6%. But for attendees who did not have a conflict, that figure rose to 3.8%.

After taking other effects into account, the authors found that meeting attendees cited liked papers 52% more often when they could see them in person than when they couldn't. That's "pretty sizable", Teplitskiy says.

The analysis also found a similar benefit for citations of non-liked papers, an effect that the authors call serendipitous diffusion of information. Serendipitous diffusion accounted for nearly 22% of the overall dissemination of information brought about by presentations at the conferences.

Serendipitous spread

That was a surprise, Teplitskiy says. The team had assumed at the start of the project that serendipitously seeing a talk at a conference would have a small effect on later citation, if the effect could be measured at all. That the serendipitous spread of ideas can be measured and quantified is important,

he says, because studying how ideas flow within communities is challenging. The study findings suggest that scheduling conflicts could be a useful tool to analyse this information flow.



Organizing a virtual conference changed the way we think about academic exchange

The work adds to evidence that attending conferences in person is important for scientists, says Ina Ganguli, an economist at the University of Massachusetts Amherst. “We know that face-to-face interactions in settings like conferences matter,” she says. “It would be interesting to quantify how much knowledge diffusion and exchange is due to being at a conference in person or whether you can get similar effects in virtual formats.”

The work is important, because “the authors found a laboratory to study the effect of something unrelated to quality on a paper’s citations”, says economist Edward Van Wesep at the University of Colorado Boulder. “We all understand that presenting our work helps it get cited, but it’s hard to prove that this is true,” he adds.

Scientists pour time and money into organizing and attending conferences, Ganguli says, “so having evidence of the benefits like those shown in this paper, is important”.

Nature **618**, 15-16 (2023)

doi: https://doi.org/10.1038/d41586-023-01604-x

References

1. Teplitskiy, M., Park, S., Thompson, N. & Karger, D. Preprint at https://arxiv.org/abs/2209.01175 (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01604-x

- NEWS
- 23 May 2023

Does the roar of rocket launches harm wildlife? These scientists seek answers

With rocket lift-offs set to increase drastically, a team will monitor the effects of noise pollution at a California spaceport.

- Nicola Jones



A bird takes flight as the private spaceflight firm SpaceX launches a rocket from Vandenberg US Space Force Base in Lompoc, California.Credit: Patrick T. Fallon/AFP Via Getty

Rocket launches are so thunderously loud that, even from kilometres away, they can feel like being at the front of the stage at a rock concert. That, plus the fact that launch numbers are booming, has prompted a physicist and biologist to team up to study the effects of noise pollution on wildlife — particularly endangered wildlife — at one of the busiest spaceports in the world.

“You can feel it in your chest — it’s like being in a vehicle that has really big speakers, and you can just feel the vibrations coming through your body,” says one of the researchers, Lucas Hall, a wildlife ecologist at California State University in Bakersfield. Hall remembers watching a launch at California’s Vandenberg Space Force Base from a safe distance. “It’s absolutely nuts.”



2022 was a record year for space launches

Hall, physical acoustician Kent Gee at Brigham Young University in Provo, Utah, and their team have close to US$1 million in funding from the US Army Corps of Engineers over 3 years to measure the soundscape and monitor a host of endangered and threatened species living near the Vandenberg base. The project, which also includes Darryl York,

conservation branch chiefat Vandenberg, was presented on 8 May at the Acoustical Society of America annual meeting in Chicago, Illinois. The space base initiated the project, Gee says, because of the threatened species nearby, in the coastal area around Santa Barbara, California. “This is responsible stewardship, in my opinion,” he says.

Vandenberg has, for decades, sent up about 5–15 rockets per year, but with satellite launches from private companies such as SpaceX increasing, the number is set to rise sharply — to 50–100 per year by 2030. That would surpass Vandenberg’s previous record of about 45 launches per year in the mid-1960s, during the heyday of the space race between the United States and the Soviet Union. Last year, there were a record 180 launches worldwide.

“If it’s happening every week, every few days, are there any implications of that?” asks Hall. “That has not been studied.”

This type of study is “incredibly important”, says acoustician Caroline Lubert at James Madison University in Harrisonburg, Virginia, who has worked with Gee but isn’t involved with this project. “It needs to be a priority. We’re only going to have more launches, not less.”

Biodiversity concerns

There have long been concerns about the environmental impact of sparked fires, and debris and noise from rocket launches — especially at prominent sites in California, Florida and Texas that are close to US biodiversity hotspots.

SpaceX’s first Starship flight on 20 April unexpectedly blew up the launch pad at the spaceport in Boca Chica, Texas, scattering pulverized concrete and sparking a wildfire in a nature reserve. The incident has triggered a lawsuit from conservation groups, arguing that the Federal Aviation Authority should have subjected the company to a more stringent environmental assessment. The region around the spaceport hosts threatened populations of turtles and birds, including piping plovers (*Charadrius*

melodus), whose numbers are declining, and Northern Aplomado falcons (*Falco femoralis*), which are endangered in Texas.



Piping plovers (*Charadrius melodus*) are among the birds that inhabit the area around a spaceport in Boca Chica, Texas, where a SpaceX rocket blew up last month.Credit: Vicki Jauron, Babylon and Beyond Photography/Getty

Sharon Wilcox, senior Texas representative for the conservation group Defenders of Wildlife, which is based in Washington DC, says the organization has "considerable concerns about the impacts of sound, concussion and vibration resulting from these launches".

Although it is clear that rocket launches are extraordinarily loud, the details aren't always predictable or well understood. For instance, Gee measured the noise from last year's launch of the uncrewed Artemis I mission to the Moon at 127–136 decibels from several kilometres away[1]. The sound's intensity was nearly two orders of magnitude greater than predicted by

models used in environmental assessments, suggesting a need to revise those models.

It's not just the volume of the launch noise that's important for nearby creatures, Gee adds, but also the distribution of sound frequencies. "We have very little information about how animals perceive these sounds," he says.

Noise tolerance

Chronic noise pollution — from cities, cars or boats, for example — is known to increase stress levels for animals ranging from whales[2] to humans. It can also have an impact on birdsong[3] and animal behaviour. But the impacts of recurring, extremely loud events such as rocket launches are not well known.



Lift off! Artemis Moon rocket launch kicks off new era of human exploration

The team will be measuring the soundscape close to Vandenberg's many launch sites and, as a control, at locations tens of kilometres away. Animals in these habitats include the western snowy plover (*Charadrius alexandrinus nivosus*), the California least tern (*Sterna antillarum browni*), the California red-legged frog (*Rana draytonii*) and the tidewater goby (*Eucyclogobius newberryi*). The team has weekly data on plover and tern numbers stretching back to the 1990s, collected by the US Fish and Wildlife Service.

Cameras will capture how animals react to rocket-launch sounds: for example, whether birds abandon their nests or change their foraging or mating behaviour. Audio monitors will pick up whether they alter their songs in response to the noise, in the same way that people yell after loud noise exposure. The birds will have some resilience, Hall says. “But at some point, there’s going to be a threshold where that resilience is overcome.”

Although Hall and Gee have secured funding for only three years, they hope to continue the work for more than a decade.

Spaceports already dampen launch noise by dumping massive amounts of water onto the launch pad to absorb sound energy. The study will help to determine whether anything needs to change to protect wildlife, such as avoiding launches during certain sensitive times such as breeding seasons, or changing the shape or size of the fire trench designed to divert rocket fumes, heat and noise. “We have to focus on whether we’re trying to reduce noise, redirect noise or change the frequency of the noise,” Lubert says.

Nature **618**, 16-17 (2023)

doi: https://doi.org/10.1038/d41586-023-01713-7

References

1. Gee, K. L. *et al. JASA Express Lett.* 3, 023601 (2023).
2. Rolland, R. M. *et al. Proc. R. Soc. B* **279**, 2363–2368 (2012).
3. De Framond, L. & Brumm, H. *Proc. R. Soc. B.* **289**, 20220906 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01713-7

- NEWS
- 23 May 2023

Game-changing obesity drugs go mainstream: what scientists are learning

Studies tackle who’s most likely to lose weight on the new generation of anti-obesity medications.

- Mariana Lenharo



US demand for weight-loss drug Wegovy has been so strong that its maker has effectively limited the number of people who can start taking the

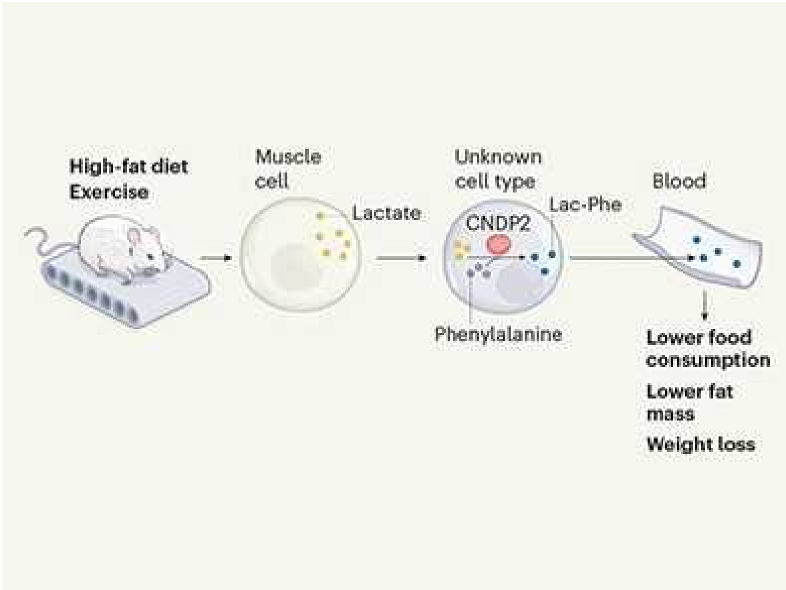
drug.Credit: EPA-EFE/Shutterstock

A new generation of drugs is revolutionizing the treatment of obesity and astonishing researchers with their potency. The drug semaglutide, for example, enabled one-third of clinical-trial participants to shed at least 20% of their body weight[1]. Tirzepatide, a competing therapy, achieved similar results in more than half of study participants[2].

Despite their high effectiveness, researchers are learning that these drugs are not necessarily the solution for everyone living with obesity. “Everybody wants to try them but not everyone responds to them,” says Andres Acosta, an obesity specialist at Mayo Clinic in Rochester, Minnesota. Now that semaglutide and tirzepatide have been around for some time, health-care providers are beginning to identify who is likely to benefit the most from them — and to recognize challenges in the drugs’ use.

Hungry brain or hungry gut?

The drugs act by mimicking a hormone called glucagon-like peptide 1 (GLP-1), which is associated with appetite regulation. In 2021, the US Food and Drug Administration approved semaglutide, under the trade name Wegovy, for treating obesity. And although the agency approved tirzepatide, marketed as Mounjaro, in 2022 for treatment of only type 2 diabetes, the drug is often prescribed as a weight-loss agent.



[Exercise molecule burns away hunger](#)

Specialists have observed that people's response to medication depends on the underlying cause of their obesity. Such observations have prompted Acosta and his colleagues to develop a method to tailor the medications to each individual[3]. The team classified people with obesity into four subtypes: those who need to eat more to reach fullness (which the researchers called hungry brain); those who reach fullness with a regular-sized meal, but feel hungry again soon (hungry gut); those who eat to cope with emotions (emotional hunger); and those with a relatively slow metabolism (slow burn).

People with a hungry gut, who tend to get hungry in between meals, seemed to respond best to the new drugs, according to anecdotal observations by Acosta and his team. Acosta says that the reason is not clear, but that he has a hypothesis. "We think that these patients have low levels of GLP-1 hormones and that's why they have obesity. And when you give them the replacement, the GLP-1 analogues, they do extremely well."

At Mayo Clinic, practitioners have used these criteria to guide treatment with an older GLP-1-mimicking drug called liraglutide. A study led by Acosta found that after using liraglutide for a year, participants with hungry-gut obesity lost twice as much weight as those from a general population of people with obesity[3].

Full up

Beverly Tchang, an endocrinologist at Weill Cornell Medicine in New York City, says that Acosta's team was the first to subdivide obesity into these categories. Her own anecdotal experience leads her to agree that people with the hungry-gut phenotype tend to do well when taking the GLP-1-mimicking drugs. Data from other groups also support Acosta's hypothesis[4]. "We do know that semaglutide slows down the gut and does improve measures of fullness," she says.

Because of their weekly regimen, the drugs are also benefiting people with erratic schedules and work shifts, notes Peminda Cabandugama, an endocrinologist at the Cleveland Clinic in Ohio and a spokesperson for The Obesity Society. Some of the other medications have to be taken at a certain time of day, he says, and "that can be difficult" for those with irregular sleep schedules.

Lifelong regimen?

Researchers hope to learn not only who is most likely to benefit but also how long people should take these medications. A clinical trial[5] showed that participants who stopped taking semaglutide after 20 weeks regained most of the weight they had initially lost, whereas those who remained on the medication continued to lose weight. Acosta says that obesity is now understood as a chronic disease and these medications are being recommended for the long term. "Now, do we know what long term means? I think nobody knows and I hope it's not for life. But those are studies that need to be done."



Obesity: The fat advantage

Another challenge, according to Tchang, is that these drugs are so efficient that practitioners are starting to discuss how much weight loss is too much. Weight loss leads to a reduction in both fat and muscle, she says. And muscle loss raises the risk of cardiovascular disease, osteoporosis and other disorders — an especially problematic outcome for older people and those with pre-existing cardiovascular conditions. Such populations are subject to what's called the obesity paradox, a correlation between weight loss and a higher rate of death.

When treating these at-risk people with the new drugs, providers are beginning to target the resolution of obesity-related problems, such as sleep apnoea, fatty liver disease and type 2 diabetes, rather than aiming for a marked weight reduction. To achieve that balance, specialists such as Tchang are increasingly exploring the possibility of prescribing lower doses.

"We want to find a place where we are treating a person for health, to prevent disease and to improve quality of life," Tchang says. "It's not necessarily about getting the most weight loss out of them."

Nature **618**, 17-18 (2023)

doi: https://doi.org/10.1038/d41586-023-01712-8

References

1. Wilding, J. P. H. *et al. N. Engl. J. Med.* **384**, 989–1002 (2021).
2. Jastreboff, A. M. *et al. N. Engl. J. Med.* **387**, 205–216 (2022).
3. Acosta, A. *et al. Obesity* **29**, 662–671 (2021).
4. Jensterle, M. *et al. Endocr. Abstr.* **81**, OC5.2 (2022).
5. Rubino, D. *et al. JAMA* **325**, 1414–1425 (2021).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01712-8

- NEWS
- 24 May 2023

Brain–spine interface allows paralysed man to walk using his thoughts

The device provides a connection between the brain and spinal cord, allowing thought to control movement.

- Dyani Lewis



Gert-Jan Oskam walks with the assistance of a crutch. A spinal-cord injury paralysed him below the neck 12 years ago, but a new device has given him

mobility.Credit: CHUV/Gilles Weber

Twelve years ago, a cycling accident left Gert-Jan Oskam, now 40, with paralysed legs and partially paralysed arms, after his spinal cord was damaged in his neck. But these days, Oskam is back on his feet and walking, thanks to a device that creates a ‘digital bridge’ between his brain and the nerves below his injury[1].

The implant has been life-changing, says Oskam. “Last week, there was something that needed to be painted and there was nobody to help me. So I took the walker and the paint, and I did it myself while I was standing,” he says.

The device — called a brain–spine interface — builds on previous work[2] by Grégoire Courtine, a neuroscientist at the Swiss Federal Institute of Technology in Lausanne and his colleagues. In 2018, they demonstrated that, when combined with intensive training, technology that stimulates the lower spine with electrical pulses can help people with spinal-cord injuries to walk again.

Oskam was one of the participants in that trial, but after three years, his improvements had plateaued. The new system makes use of the spinal implant that Oskam already has, and pairs it with two disc-shaped implants inserted into his skull so that two 64-electrode grids rest against the membrane covering the brain.

When Oskam thinks about walking, the skull implants detect electrical activity in the cortex, the outer layer of the brain. This signal is wirelessly transmitted and decoded by a computer that Oskam wears in a backpack, which then transmits the information to the spinal pulse generator.

The previous device “was more of a pre-programmed stimulation” that generated robotic stepping movements, says Courtine. “Now, it’s completely different, because Gert-Jan has full control over the parameter of stimulation, which means that he can stop, he can walk, he can climb up staircases.”

“The stimulation before was controlling me and now I am controlling stimulation by my thought,” says Oskam. “When I decide to make a step, the simulation will kick in, as soon as I think about it.”

Enhanced rehabilitation

After around 40 rehabilitation sessions using the brain–spine interface, Oskam had regained the ability to voluntarily move his legs and feet. That type of voluntary movement was not possible after spinal stimulation alone, and suggests that the training sessions with the new device prompted further recovery in nerve cells that were not completely severed during his injury. Oskam can also walk short distances without the device if he uses crutches.

Bruce Harland, a neuroscientist at the University of Auckland in New Zealand, says that this continued improvement in spinal function is great news for anyone with a spinal-cord injury, “because even if it’s a longer-term chronic injury, there’s still a few different ways that healing could happen”.

“It’s certainly a huge jump” towards improved function for people with spinal-cord injuries, says neuroscientist Anna Leonard at the University of Adelaide in Australia. And, she says, there is still room for other interventions — such as stem cells — to improve outcomes further. She adds that although the brain–spine interface restores walking, other functions such as bladder and bowel control are not targeted by the device. “So, there’s certainly still room for other areas of research that could help progress improvements in outcomes for these other sort of realms,” she says.

Antonio Lauto, a biomedical engineer at Western Sydney University, Australia, says that less-invasive devices would be ideal. One of Oskam’s skull implants was removed after about five months because of an infection. Nevertheless, Jocelyne Bloch, the neurosurgeon at the Swiss Federal Institute of Technology who implanted the device, says that the risks involved are small compared with the benefits. “There is always a bit of risk of infections or risk of haemorrhage, but they are so small that it’s worth the risk,” she says.

Courtine's team is currently recruiting three people to see whether a similar device can restore arm movements.

Nature **618**, 18 (2023)

doi: https://doi.org/10.1038/d41586-023-01728-0

References

1. Lorach, H. *et al. Nature* https://doi.org/10.1038/s41586-023-06094-5 (2023).

2. Wagner, F. B. *et al. Nature* **563**, 65–71 (2018).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01728-0

- NEWS
- 18 May 2023

JWST spots biggest water plume yet spewing from a moon of Saturn

The huge watery cloud spurting from Enceladus could carry the ingredients for life farther into space than previously known.

- Alexandra Witze



The Cassini spacecraft swoops through plumes issuing from Saturn’s moon Enceladus (artist’s impression).Credit: NASA/JPL-Caltech

The James Webb Space Telescope (JWST) has spotted Saturn’s moon Enceladus spraying out a huge plume of water vapour, much bigger than any previously seen there. This enormous cloud might contain the chemical ingredients of life, escaping from beneath the moon’s icy surface.

In 2005, a NASA spacecraft called Cassini discovered icy particles squirting from Enceladus's subsurface ocean through cracks in the moon's surface. But JWST shows that material is spraying much farther than previously thought — many times deeper into space than the size of Enceladus itself.

"It's immense," said Sara Faggi, a planetary astronomer at NASA's Goddard Space Flight Center in Greenbelt, Maryland, on 17 May at a conference at the Space Telescope Science Institute in Baltimore, Maryland. She declined to provide further details, citing a scientific paper that will be published soon.

Rare ocean world

Enceladus excites astrobiologists because it is one of the few 'ocean worlds' in the Solar System, making it one of the best places to look for extraterrestrial life. The salty ocean that lies beneath Enceladus's outer covering of ice is a possible haven for living organisms, which could be sustained by chemical energy at hydrothermal vents on the ocean floor.

The material that squirts out of Enceladus, primarily through fractures known as tiger stripes around the moon's south pole, is a direct link to that potential extraterrestrial ecosystem. The plumes seen by Cassini contained silica particles that were probably carried up from the sea floor by churning fluids[1]. Cassini flew many times through Enceladus's plumes, measuring ice grains and life-friendly chemicals such as methane, carbon dioxide and ammonia.

But it took JWST, a telescope located 1.5 million kilometres from Earth, to discover something that Cassini could not see from its ringside seat. Whereas Cassini could spot ice grains that do not travel far from the surface, JWST has a wider perspective and sensitive instruments that can capture faint gas signals around Enceladus.

Enceladus at a glance

On 9 November 2022, JWST peeked briefly at Enceladus. Just 4.5 minutes' worth of data revealed the enormous, very cold plume of water vapour. The

forthcoming paper will quantify how much water is spraying out and its temperature, Faggi said. But the plume is likely to be of low density, more like a diffuse, cold cloud than a damp spray. That's not great news for anyone looking to grab samples from the plume and hoping to find life, because the signs of life might be too sparse to detect[2]. Ice grains seen by Cassini much closer to Enceladus are more likely to have high concentrations of organic particles, says Shannon MacKenzie, a planetary scientist at the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland.



Scientists made a new kind of ice that might exist on distant moons

JWST also analysed the spectrum of sunlight reflecting off Enceladus and found evidence of many chemicals, including water and possibly other compounds that could hint at geological or biological activity in the moon's ocean. "We have many more surprises," Faggi said.

Researchers are already planning how to follow up on the discovery. Last week, JWST organizers released a list of the observations to be taken in the telescope's second round of operations — and it includes another project to study Enceladus. That work will look at Enceladus for six times longer than the first JWST study, and will aim to find chemical compounds associated with habitability, such as organic compounds and hydrogen peroxide. "The new observation will give us our best shot yet at searching for habitability

indicators on the surface,” says project lead Christopher Glein, a geochemist at the Southwest Research Institute in San Antonio, Texas.

Moon-scouting snake robot

JWST’s findings provide more grist for a possible NASA mission to Enceladus to search for signs of life there. Proposals under consideration include an ‘orbilander’ mission that would orbit the moon for a year and a half before landing at its south pole. Another proposal calls for the development of an autonomous snake robot that could slither beneath Enceladus’s ice to explore the ocean.

Other icy moons in the Solar System are also getting attention from JWST. At the conference, Geronimo Villanueva, a planetary scientist at Goddard, reported that the telescope had detected carbon dioxide on Jupiter’s moon Europa. That excites scientists because carbon and oxygen are key building blocks for life on Earth. NASA is launching a mission to Europa next year that will explore that ocean world in more detail. “This is definitely a new era in the exploration of the Solar System,” Villanueva said.

Nature **618**, 19 (2023)

doi: https://doi.org/10.1038/d41586-023-01666-x

References

1. Schoenfeld, A. M. *et al. Comm. Earth Environ.* **4**, 28 (2023).

2. Affholder, A., Guyot, F., Sauterey, B., Ferrière, R. & Mazevet, S. *Planet. Sci. J.* **3**, 270 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01666-x

- NEWS EXPLAINER
- 19 May 2023

When will global warming actually hit the landmark 1.5 ºC limit?

The planet is on track to reach the 1.5 ºC average by the 2030s — although a new report suggests a single year will probably cross the line much sooner.

- Nicola Jones



Temperatures soared in India in April: heatwaves become more likely as the planet heats up. Credit: Sudipta Das/NurPhoto via Getty

There's a 66% chance that the annual global average temperature will hit 1.5 ºC above pre-industrial temperatures at some time in the next five years,

according to a World Meteorological Organization report released on 17 May. Reaching 1.5 ºC of warming in a single year will be a landmark moment for the planet, which in 2022 was about 1.15 ºC warmer than in pre-industrial times. But it's not quite the milestone most people mean when they talk about 1.5 ºC of warming — for that, we probably have about a decade to go.

The famous 1.5 ºC figure, widely quoted as the desired 'maximum' for planetary warming, stems from the 2015 United Nations Paris agreement on climate change. This treaty declared the goal of keeping the global average temperature well below 2 ºC above pre-industrial levels, with a preferred limit of 1.5 ºC.

The Paris agreement, however, refers to a sustained planetary average of 1.5 ºC warming — not just the average for a single year, which alone could be anomalously hotter or cooler than the longer-term average. The Paris agreement didn't specify exactly what was meant by 1.5 ºC of warming, but the most recent report of the first working group of the Intergovernmental Panel on Climate Change (IPCC), published in 2021, clarifies that it means the midpoint of the first 20-year period when the average global surface air temperature is 1.5 ºC warmer than the 1850–1900 average.

In 2018, an IPCC special report on 1.5 ºC of warming estimated that the world would probably hit the 1.5 ºC threshold at some stage between 2030 and 2052. By 2021, using a different methodology, that had been pinned down to the early 2030s. "The time frame is getting closer and closer," says geographer William Solecki at City University of New York, an author on the IPCC special report.

A massive, two-year 'global stocktake' of progress on the Paris agreement's goals is winding up now, and will be presented at the next UN Framework Convention on Climate Change meeting (COP28), which will start on 30 November. So far, the stocktake has found that things aren't going well. For a 50% chance of limiting warming to 1.5 ºC, a stocktake meeting report notes, global greenhouse-gas emissions need to peak before 2025; this hasn't happened yet, and national emissions commitments aren't sufficient to keep the planet within the target.

The lower the better

The 1.5 ºC number was chosen in an attempt to limit the severity of the impacts of warming, taking into account factors such as food security and extreme weather events. However, IPCC experts stressed that 1.5 ºC shouldn't be seen as a "guardrail" below which everything would be fine, and noted that whatever temperature the world's warming peaks at, the lower it is, the better. "Obviously there's a continuum," says Solecki. "The higher the temperature, the worse the outcome."

The 2018 IPCC report on 1.5 ºC of warming notes that effects of reaching this threshold could include: extreme hot days in mid-latitudes that are 3 ºC warmer than in pre-industrial times; sea-level rise of up to three-quarters of a metre by 2100; the loss of more than half of the viable habitat for 8% of plants and 4% of vertebrates; and a decrease in annual global fisheries catches of 1.5 million tonnes.

The report also notes that because global warming is uneven, more than one-fifth of the world's population currently live in regions that have already exceeded 1.5 ºC of warming in at least one season.

More important than when Earth will hit 1.5 ºC of warming is what amount of warming the planet will peak at, and when that will happen. "With every tenth of a degree above 2 ºC, you're looking at more-sustained, more-systemic impacts," says Solecki.

Those numbers won't be apparent for decades. According to the IPCC's 2021 projections of global temperature under different emissions scenarios, peak temperature could be anything from 1.6 ºC in around 2050 (if the globe hits net zero emissions by then), dropping to 1.4 ºC by 2100; to, with emissions still climbing, 4.4 ºC at 2100, with the peak still to come.

The next few years could bring an anomalously high blip in annual temperatures compared with the longer-term average thanks to an expected El Niño event — a natural climate pattern that brings warmer temperatures to the eastern Pacific Ocean and that tends to warm the planet as a whole. In April, Carbon Brief, a website that reports on climate matters, estimated that 2023 was shaping up to be one of the six hottest years on record, most likely

the fourth hottest. And in April, the global ocean spiked at the hottest temperature since records began.

Nature **618**, 20 (2023)

doi: https://doi.org/10.1038/d41586-023-01702-w

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01702-w

- NEWS FEATURE
- 31 May 2023

‘Almost magical’: chemists can now move single atoms in and out of a molecule’s core

An explosion of skeletal editing methods to insert, delete or swap individual atoms in molecular backbones could accelerate drug discovery.

- Mark Peplow

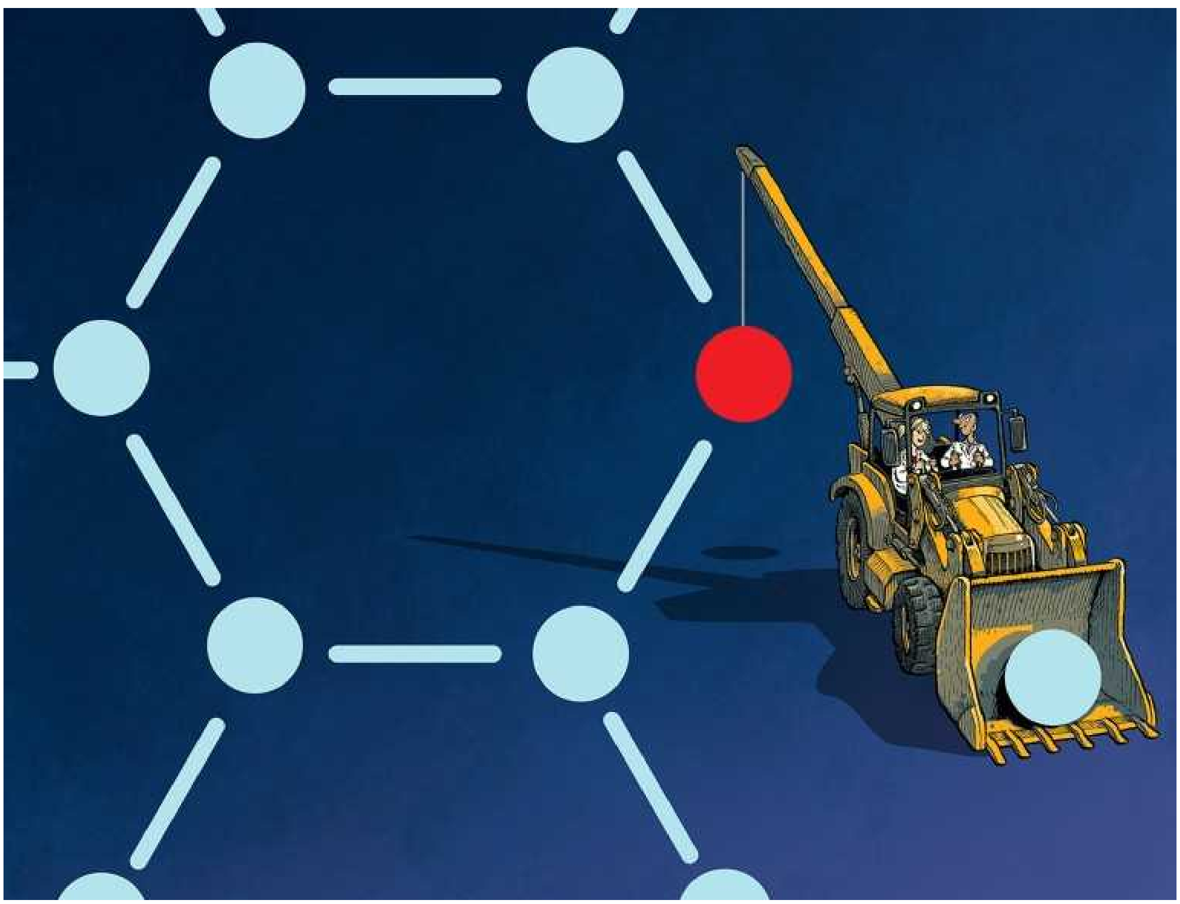
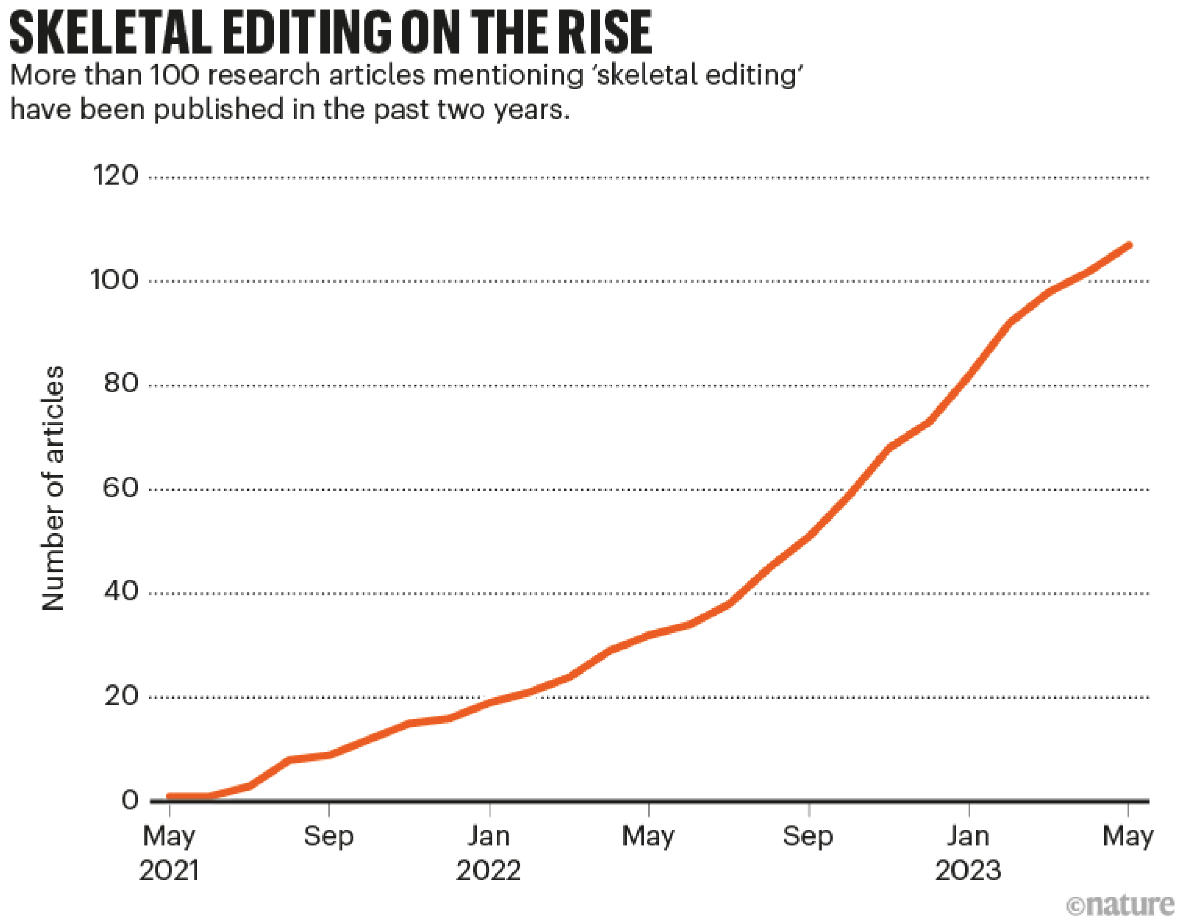


Illustration by David Parkins

Half a decade ago, chemist Mark Levin was a postdoc looking for a visionary project that could change his field. He found inspiration in a set of published wish lists from pharmaceutical-industry scientists who were looking for ways to transform medicinal chemistry[1,2]. Among their dreams, one concept stood out: the ability to precisely edit a molecule by deleting, adding or swapping single atoms in its core.

This sort of molecular surgery could dramatically speed up drug discovery — and might altogether revolutionize how organic chemists design molecules. One 2018 review called it a ‘moonshot’ concept. Levin was hooked.

Now head of a team at the University of Chicago in Illinois, Levin is among a cadre of chemists pioneering these techniques, aiming to more efficiently forge new drugs, polymers and biological molecules such as peptides. In the past two years, more than 100 papers on the technique — known as skeletal editing — have been published, demonstrating its potential (see ‘Skeletal editing on the rise’). “There’s a tremendous amount of buzz right now around this topic,” says Danielle Schultz, director of discovery-process chemistry at pharmaceutical company Merck in Kenilworth, New Jersey.



Source: *Nature* analysis using Digital Science Dimensions database.

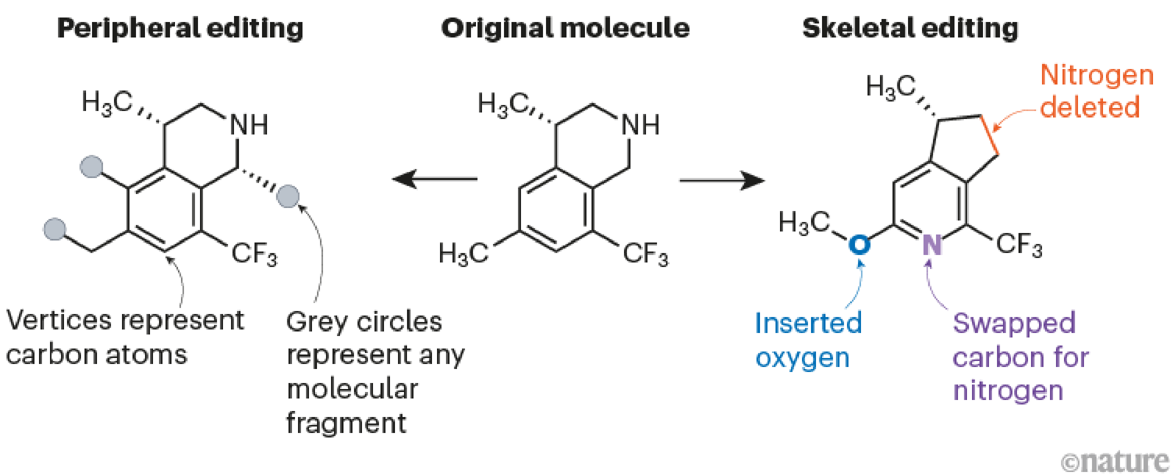
To get a sense of the challenge, consider that the small, carbon-based molecules that make up most of the world's drugs typically contain fewer than 100 atoms, and are assembled piece by piece in a series of chemical reactions. Some connect large sections of the molecule's skeleton; others decorate that skeleton with clusters of atoms to create the final product. But few methods can reliably tweak a molecule's core skeleton once it has been assembled. It's a little like clipping together a house from Lego bricks: remodelling the exterior is trivial, but inserting a brick into the middle of a completed wall can't be done without taking the house apart.

For organic chemists, the idea of being able to swap an atom in a molecule's skeleton holds an intrinsic fascination. "It's almost magical that these changes are now possible," says Richmond Sarpong at the University of California, Berkeley, a leading light in skeletal editing.

But there's also a very practical purpose. Drug discovery involves first finding a promising molecule, and then making hundreds of slightly different versions to try to improve potency or reduce toxicity. It's relatively easy to change atomic groups on a molecule's periphery to make variants. To edit the core, however, researchers typically must return to the start of their synthesis and make the modified skeleton from scratch. This is expensive, time-consuming and, in practice, heavily limits the variety of designs that drug firms screen and test. Reliable skeletal editing could massively speed up the process (see 'The emerging chemistry of skeletal editing').

THE EMERGING CHEMISTRY OF SKELETAL EDITING

Chemists can easily add or change fragments on a molecule's periphery (left). Skeletal editing — changing the core skeleton of a molecule — is much more difficult, but chemists can now use reactions to insert, delete or swap atoms in these core areas (right).



Source: Ref. 9

At this stage, many of these methods only work on specific molecules, or edit them inefficiently. Researchers are excited but also wary of overhyping a young field.

Still, chemists at drug makers Pfizer and Merck are already testing various skeletal editing reactions. "We recognize this could be transformative," says David Blakemore, head of synthesis, inflammation, immunology and anti-infectives chemistry at Pfizer in Groton, Connecticut. "It's fairly early days,

but I don’t think we’re a long way away from being able to use some of these methods.”

Skeleton keys

To the uninitiated, organic chemistry looks like a baffling parade of hieroglyphs — a blizzard of zigzags and hexagons whirling across the page. Yet these diagrams are a visual lexicon, rich in information about the molecules they represent. Each vertex in a shape represents a carbon atom (accompanied by hydrogen atoms that are not shown), while the lines between them are chemical bonds. Then there’s a liberal seasoning of letters, representing atoms such as oxygen, nitrogen or sulfur, peppered across the molecule’s skeleton and periphery.

When chemists want to make a particular molecule, they start by sketching its structure in this way. Then they progressively carve this drawing into smaller pieces by erasing bonds that could be forged by reliable reactions, leaving simpler fragments that can be bought from chemical suppliers, or brewed up from scratch.



Mark Levin, a pioneer of skeletal editing, working at a glovebox (a sealed container for hazardous or air-sensitive chemistry).Credit: Jean Lachat Photography

Adding a reliable set of skeletal editing reactions to the chemists' toolkit could provide a completely new set of disconnections, enabling more efficient syntheses and ways to make previously inaccessible compounds. "It's a very clear departure from how we used to think about synthesis," says organic chemist Junqi Li at Iowa State University in Ames.

The pay-offs could be enormous. New kinds of chemical reaction that synthesize molecules have transformed drug discovery. The [2005 Nobel Prize in Chemistry] went, in part, to researchers who had developed efficient catalysts for metathesis reactions in the 1990s. These weld together chemical groups called alkenes (C=C double bonds); the method was crucial for creating the large molecular rings in a series of hepatitis C drugs, for example. The [2010 chemistry Nobel] went to researchers who, in the 1960s and 1970s, had developed palladium-catalysed cross-coupling reactions, offering another way to forge carbon–carbon bonds in drug molecules.

In the twenty-first century, there have been major advances in reactions that edit a molecule's periphery. Carbon–hydrogen (C–H) bonds are usually passive bystanders in chemistry, but the past two decades have seen a parade of methods (known generally as C–H activation) to replace hydrogen with clusters of atoms that have more useful chemical properties, called functional groups.

If skeletal editing could be applied to a broad range of molecules — adding or removing specific atoms with complete selectivity, and without damaging existing functional groups — the approach would rank alongside those past innovations, says David Rees, chief scientific officer at Astex Pharmaceuticals in Cambridge, UK. "I think it would be potentially even bigger than any of them."

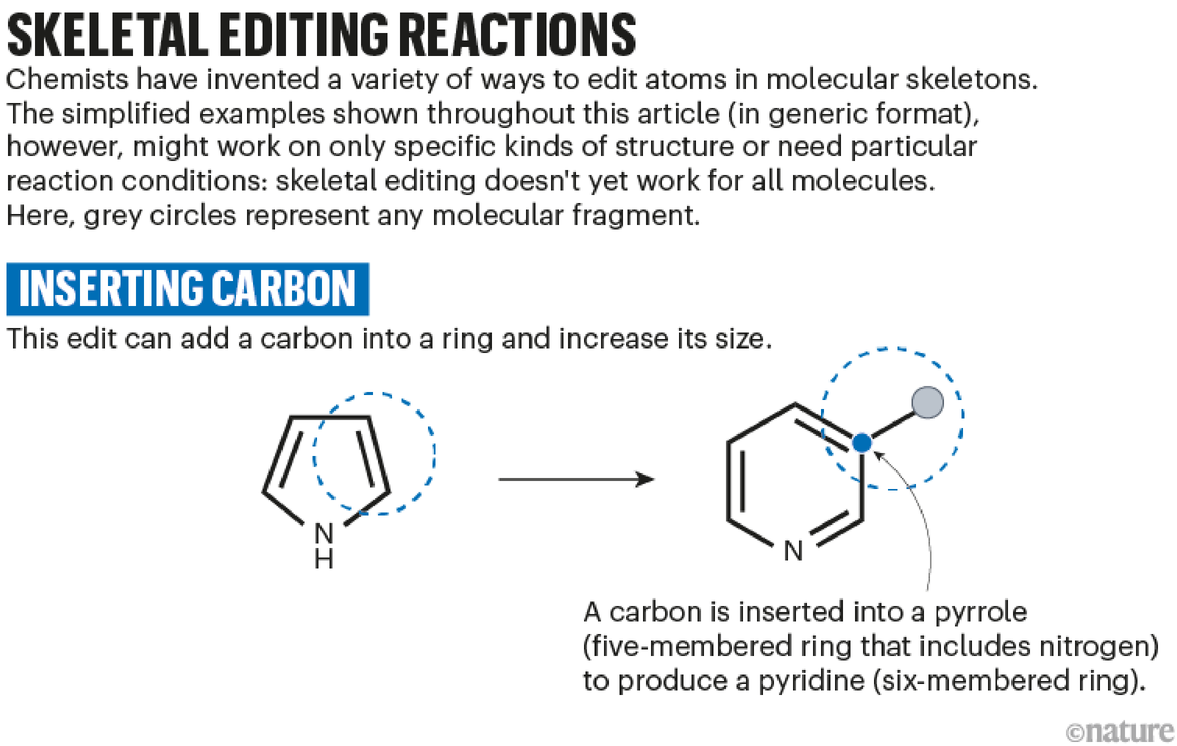
That, however, is a tough task. It's perhaps tempting to draw a parallel between skeletal editing and the gene-editing technique CRISPR that has taken biotechnology by storm. In fact, says Levin, it's a poor analogy. CRISPR only has to cope with the four bases in DNA or RNA — but for

skeletal editing to be generalizable, the editing techniques will have to work reliably on thousands of different molecules.

Meeting medicinal needs

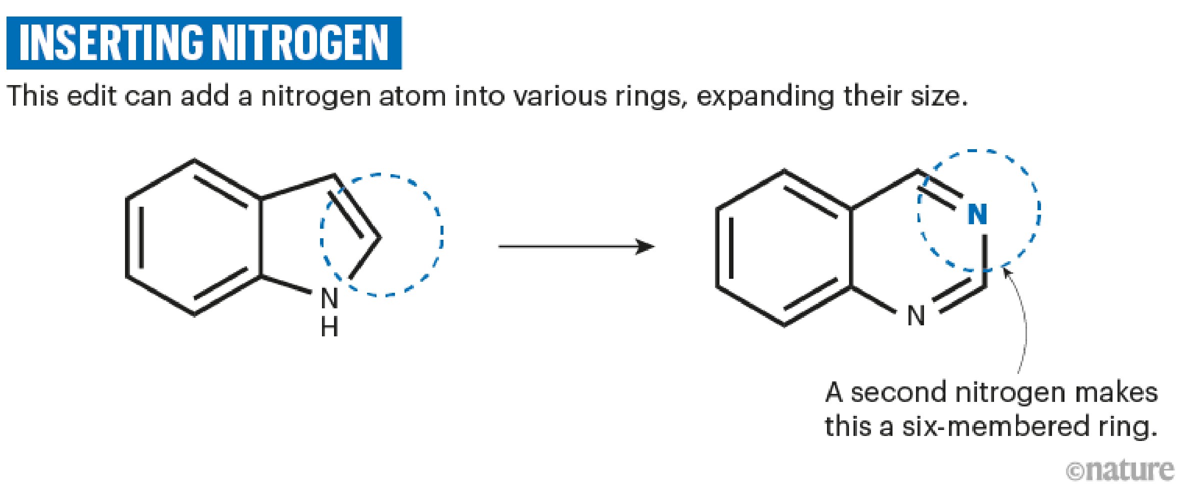
The skeletal editing reactions that chemists have unveiled are generally tailored to the needs of medicinal chemists[3]. Many are designed to edit molecular rings that contain non-carbon atoms (such as O, N or S): known as heterocycles, they are ubiquitous in medicinal chemistry. About 60% of small-molecule drugs contain a nitrogen heterocycle, for example, in part because it often helps a drug to bind well to protein targets[4]. Medicinal chemists might want to delete one atom from a heterocycle so that the ring contracts, changing its fit for a protein site, or add an atom that strengthens its binding. Alternatively, they might want to tweak the molecule to improve its solubility or reduce its toxicity.

Levin and his group, for instance, have developed a way to insert a carbon atom into a pentagon-shaped nitrogen heterocycle to create a hexagonal one (see ‘Skeletal editing reactions: inserting carbon’). In the biochemical jargon, this transforms pyrroles into pyridines[5]. “That is a really powerful reaction,” Blakemore says. Pyridines are the second-most-common heterocycle in drugs approved by the US Food and Drug Administration. However, heavily decorated pyridines are much harder to make than the corresponding pyrroles, so Levin’s reaction could offer a much simpler route to those desirable skeletons.



Source: Ref. 5; reaction details omitted.

Julia Reisenbauer, a PhD student working in organic chemist Bill Morandi’s team at the Swiss Federal Institute of Technology in Zurich, has developed a similar reaction to insert nitrogen into various rings[6,7] (see ‘Inserting nitrogen’). “For quite a lot of these products, it would have been way harder to access them with known methods,” she says.



Source: Ref. 6; reaction details omitted.

Medicinal chemists who want to optimize the effects of nitrogen atoms in their drug candidates sometimes perform 'nitrogen scans', testing the bioactivity of a series of analogues that each have a nitrogen atom in almost every conceivable position[8].

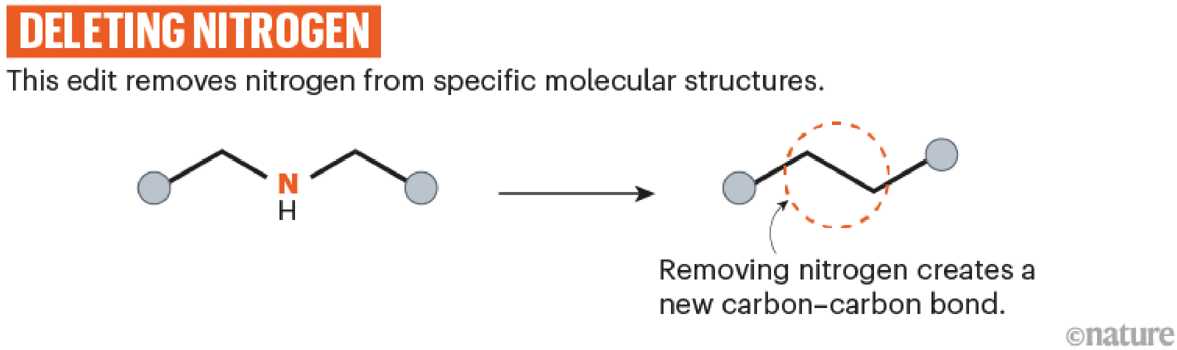
But constructing these variants is time-consuming, and can be chemically challenging: the nitrogen atoms sometimes interfere with the catalysts and chemical reagents used in syntheses, Rees says. Using skeletal editing to insert nitrogen at the end of a synthesis might circumvent that problem, and offer a practical way to perform a nitrogen scan without completely resynthesizing each analogue.



Julia Reisenbauer's nitrogen-insertion edit could help medicinal chemists to perform nitrogen scans.Credit: Marius Lutz

There's also a growing selection of deletion reactions, which either shift an atom from a molecule's skeleton to its periphery, or remove it altogether. Levin and his group have reported a reaction to snip nitrogen out of a C–N–

C group, a motif commonly found in drugs[9] (see ‘Deleting nitrogen’) — although the reagent they use doesn’t work well with bulkier molecules, and tends to attack some sensitive functional groups.

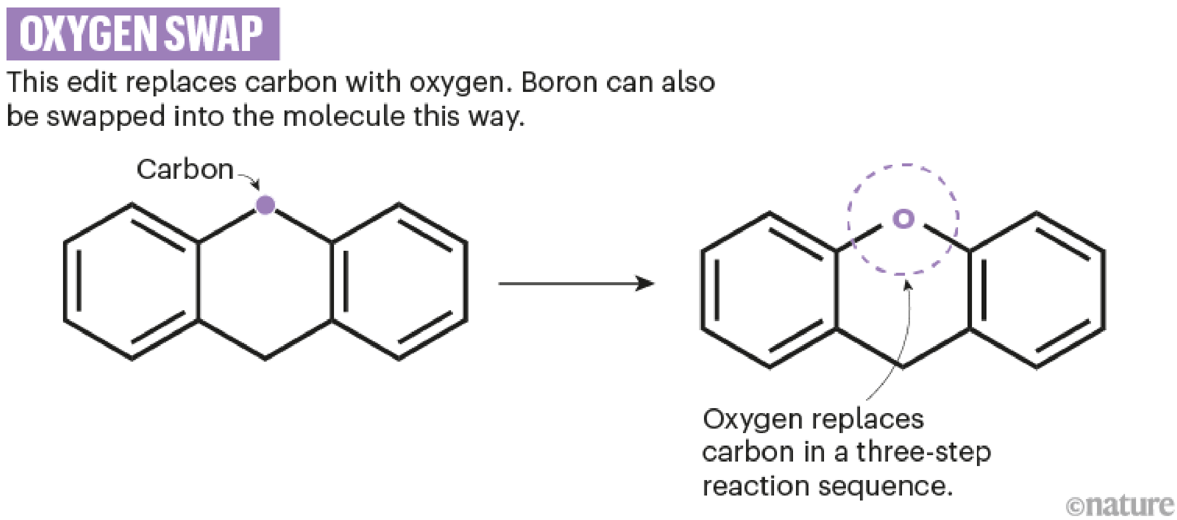


Source: Ref. 9; reaction details omitted.

Sarpong’s group, meanwhile, has come up with a method to delete a carbon that sits between two nitrogen atoms in six-membered rings called pyrimidines[10]. (The cytosine and thymine bases in DNA are both pyrimidine structures.)

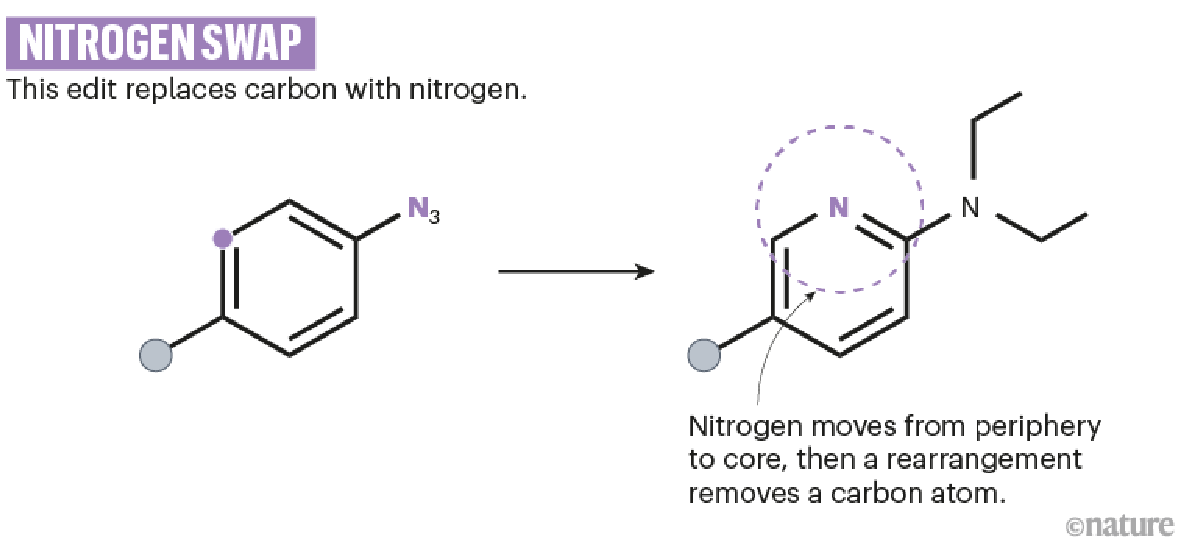
For some researchers, the most prized skeletal edits would swap one atom for another in a single step. “I’ve been obsessed with this idea of atom replacements in any skeleton,” says Levin.

Last year, Li and her colleague Quang Luu, also at Iowa State University, unveiled a reaction sequence that replaces a carbon atom with oxygen, a change that can potentially improve a molecule’s solubility in water[11] (see ‘Oxygen swap’). But so far, the method has been used only on a specific type of molecular skeleton containing a carbon atom flanked by two benzene-based groups. Li has also adapted the approach to replace the carbon atom with boron, an atom that can help drug molecules bind to sugars or proteins.



Source: Ref. 11; reaction details omitted.

At Stanford University in California, chemists Noah Burns and Sajan Patel have developed a carbon-to-nitrogen swap that is driven by blue light and oxygen[12] (see ‘Nitrogen swap’). However, it also involves a highly reactive compound called an azide that has a reputation for explosive instability.



Source: Ref. 12; reaction details omitted.

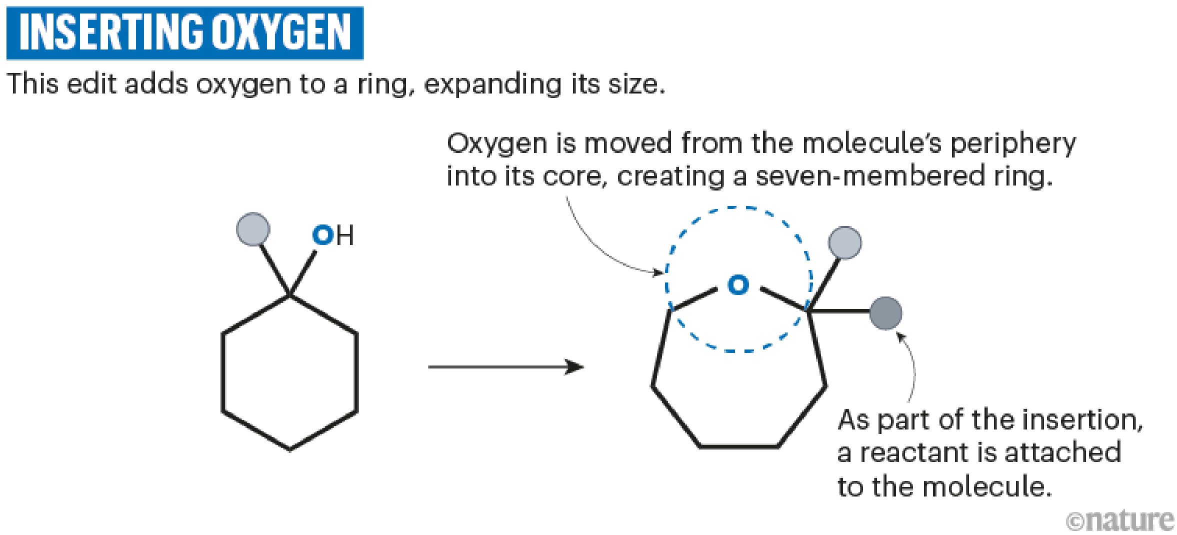
For now, “there’s still no general solution” for atom swapping, says Levin. “But we have a couple of things cooking in my lab, which aren’t ready for prime time, that are exact answers to this challenge.”

Warp-speed editing

As with other reactions, all of these innovations use reagents, catalysts or light to push or pull around atoms' shared electrons — the 'glue' that holds atoms together in the chemical bonds of organic molecules — so that new atoms can be brought in or existing ones excised. And some of the editing reactions have deep historical roots — several have enabled skeletal edits since the late nineteenth century. The Baeyer–Villiger oxidation, for example, inserts an oxygen atom; the Beckmann rearrangement inserts nitrogen, a process that every year produces millions of tonnes of caprolactam, the feedstock for nylon. (Like so many reactions in organic chemistry, these reactions bear the names of their inventors).

But these historical approaches have limited scope. They can only insert atoms next to a functional group known as a carbonyl, because they rely on its chemical reactivity to help prise open a molecule. Other skeletal editing techniques developed decades ago are rarely used, because they chew up too many functional groups in molecules or produce messy mixtures that require laborious purification.

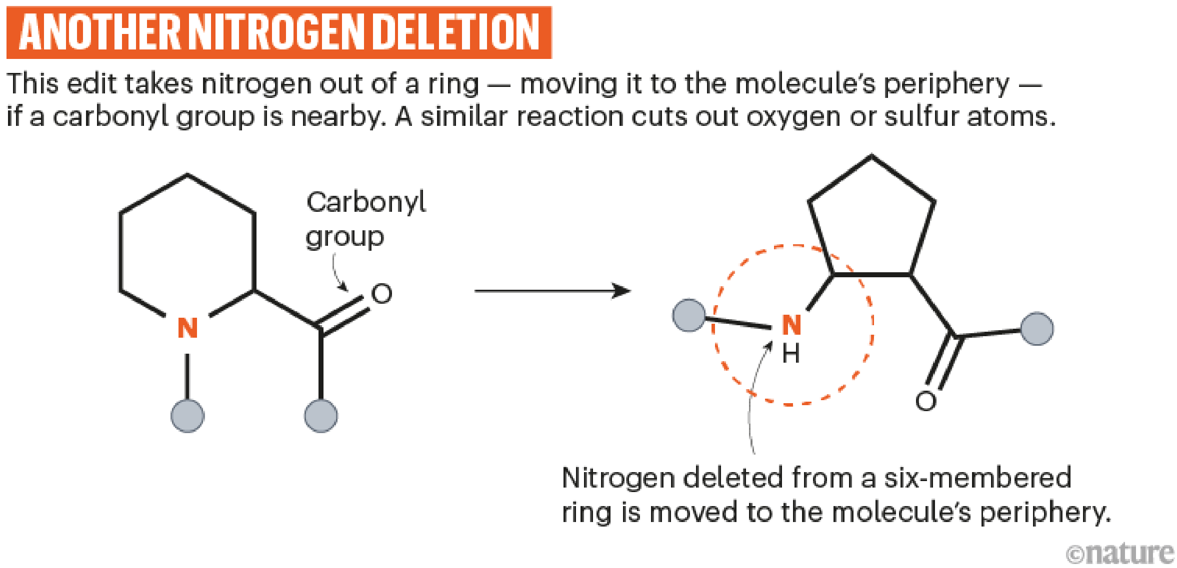
Access to modern catalysts and reagents has made improvements possible, says Reisenbauer. For instance, Levin's and Reisenbauer's atom insertions both mimic a method from the 1880s: the Ciamician–Dennstedt rearrangement. This reaction tended to damage peripheral functional groups, limiting its use, but the improved approaches use more-forgiving reagents to generate the reactive atoms needed for a clean insertion. “We're always building on the foundations and discoveries of those that came before us,” says chemist Sarah Wengryniuk at Temple University in Philadelphia, Pennsylvania, part of a team that has developed an oxygen-insertion edit[13] (see 'Inserting oxygen').



Source: Ref. 13; reaction details omitted.

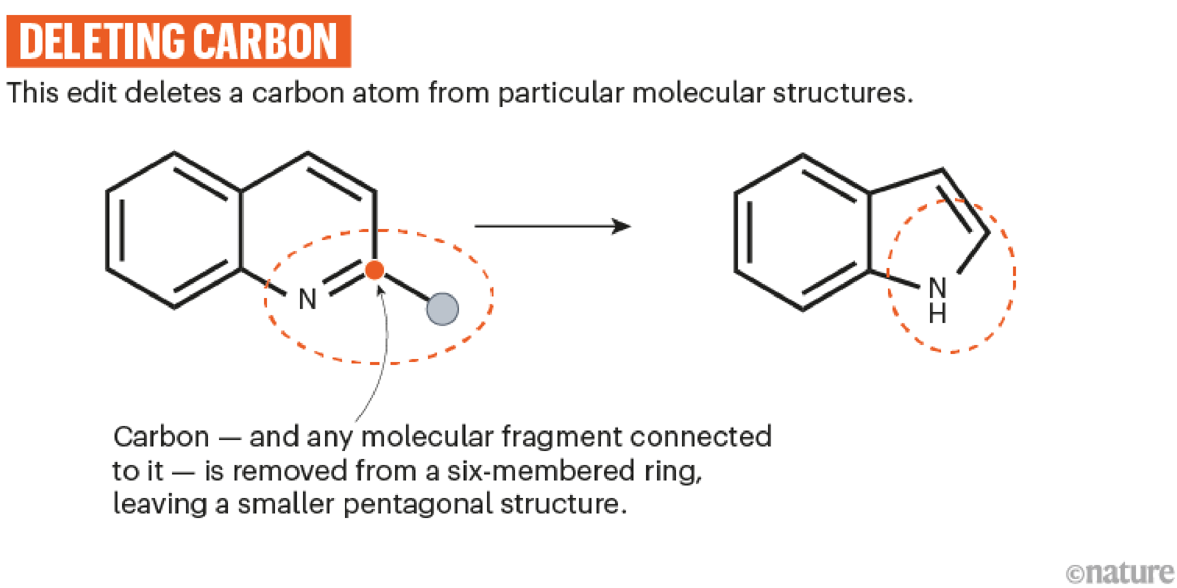
Another advance lies in the light-driven reactions that break and rejoin atomic bonds. Decades ago, these reactions relied on lamps that produced a broad spread of wavelengths, which sometimes destroyed part of a reaction's product as soon as it formed. These days, chemists use light-emitting diodes (LEDs) that emit specific wavelengths of light, which can trigger precise edits.

Sarpong has developed a light-driven reaction that cuts out N, O or S from rings, if there is a carbonyl group close to the atom being excised[14] (see 'Another nitrogen deletion'). The key was to use exactly the right wavelength of blue light to excite this carbonyl group and trigger an atom-removing rearrangement.



Source: Ref. 14; reaction details omitted.

And Levin's team, collaborating with scientists at Merck, have developed a reaction that deletes a carbon atom from some nitrogen heterocycles using violet light at a wavelength of 390 nanometres[15] (see 'Deleting carbon').



Source: Ref. 15; reaction details omitted.

Bones of contention

Skeletal editing could also assist in the earlier stages of drug discovery — helping medicinal chemists to access a broader range of molecules to screen for drug-like activity.

Chemists imagine the universe of all possible organic molecules as a territory called chemical space. It includes up to 10^{60} possible drug-like molecules, each a twinkling star of potential medicinal benefit[16]. Ideally, pharmaceutical companies' screening libraries should feature representatives from across the chemical cosmos. But, in reality, molecular structures that are easier to make tend to be over-represented in these libraries, leaving large unilluminated voids in medicinal chemical space[17].

Skeletal editing could help to push back these frontiers. By offering a one-step method to change the skeleton of a molecule, these reactions act as a wormhole in chemical space, enabling researchers to travel from one galaxy of compounds to another. This strategy is already being trialled in some pharmaceutical companies, Sarpong says (both he and Levin collaborate with Merck researchers).



Chemist Richmond Sarpong, a leading light in skeletal editing.Credit: Brandon Wright

Yet even the most enthusiastic proponents of skeletal editing concede that there are many obstacles to routine use. Some reactions are not very efficient, converting only a fraction of molecules into the desired product. (That might not matter for medicinal chemists, however, because the early stages of drug discovery often require only a tiny amount of a desired molecule.)

Another difficulty is that many edits rely on highly reactive reagents that generate unwanted by-products or might be considered too hazardous in industry. “This is my biggest reservation right now,” says Blakemore.

Sometimes, extra molecular fragments have to be pre-installed to protect vulnerable functional groups, and then removed after the reaction is complete. Although this concept is well established in chemistry, it makes the overall editing less efficient.



Sarah Wengryniuk, part of a team that has developed an oxygen-insertion edit.Credit: Courtney Ay Photography

Perhaps the biggest challenge is to make the methods more general, so that each reaction works reliably on a wide range of skeletons. Yet many of the most useful reactions in medicinal chemistry initially had only a limited scope, and there are hints that skeletal editing could likewise become more versatile and convenient.

One promising sign comes from studying how atoms and bonds shift around as reactions progress, forming fleeting intermediate forms called transition states. Some of the skeletal editing reactions share similar mechanisms, suggesting that common, and perhaps generalizable, reaction motifs are at play. Finding milder and more selective reagents to reach these kinds of intermediate might unlock a wider range of edits.

Wider horizons

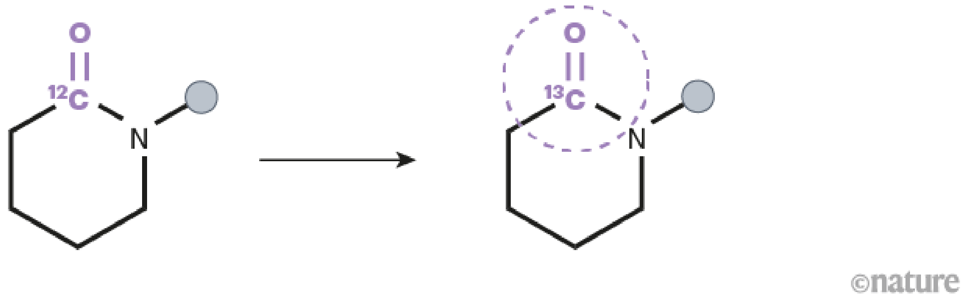
As chemists work to strengthen the foundations of skeletal editing, these techniques are already reaching areas beyond drug discovery. For instance, polymer chemists Aleksandr Zhukhovitskiy and Rachael Ditzler at the University of North Carolina at Chapel Hill have shown that similar reactions can edit the carbon-based backbones of polymers[18]. Such reactions might eventually help to recycle plastics, or make it easier to use sustainable biological molecules to produce everyday polymers that are currently made from fossil-fuel feedstocks.

Sarpong has used his light-driven nitrogen deletion to modify the structures of peptides, which are short chains of amino acids. Schultz says that a general method to do this could be incredibly useful. For instance, an oral peptide drug edited to contain non-natural amino-acid structures might be better able to withstand digestive enzymes — without losing its potency. “It could be a pretty game-changing technology,” she says.

Meanwhile, Morandi’s team has inserted unusual isotopes, such as nitrogen-15 or carbon-13, into drug molecules[7,19] (see ‘Isotope swap’). Medicinal chemists could use this strategy in experiments to track the metabolism of a drug, for example, or to understand how it interacts with its protein target.

ISOTOPE SWAP

In this edit, a carbon-12 atom is swapped for a carbon-13 isotope.



Source: Ref. 19; reaction details omitted.

It’s clear that the skeletal editors are brimming with ideas for new methods, and the field is full of early-career researchers who are competitive but also collaborative. “We want to come in and do something that people recognize as new and interesting and different,” says Levin. “I think it’s the future.”

Nature **618**, 21-24 (2023)

doi: https://doi.org/10.1038/d41586-023-01735-1

References

1. Blakemore, D. C. *et al. Nature Chem.* **10**, 383–394 (2018).

2. Campos, K. R. *et al. Science* **363**, eaat0805 (2019).

3. Jurczyk, J. *et al. Nature Synth.* **1**, 352–364 (2022).

4. Vitaku, E., Smith, D. T. & Njardarson, J. T. *J. Med. Chem.* **57**, 10257–10274 (2014).

5. Dherange, B. D., Kelly, P. Q., Liles, J. P., Sigman, M. S. & Levin, M. D. *J. Am. Chem. Soc.* **143**, 11337–11344 (2021).

6. Reisenbauer, J. C. *et al. Science* **377**, 1104–1109 (2022).

7. Finkelstein, P. *et al. Chem. Sci.* **14**, 2954–2959 (2023).

8. Pennington, L. D. & Moustakas, D. T. *J. Med. Chem.* **60**, 3552–3579 (2017).

9. Kennedy, S. H., Dherange, B. D., Berger, K. J. & Levin, M. D. *Nature* **593**, 223–227 (2021).

10. Bartholomew, G. L, Carpaneto, F. & Sarpong, R. *J. Am. Chem. Soc.* **144**, 22309–22315 (2022).

11. Luu, Q. H. & Li, J. *Chem. Sci.* **13**, 1095–1100 (2022).

12. Patel, S. C. & Burns, N. Z. *J. Am. Chem. Soc.* **144**, 17797–17802 (2022).

13. Kelley, B. T., Walters, J. C. & Wengryniuk, S. E. *Org. Lett.* **18**, 1896–1899 (2016).

14. Jurczyk, J. *et al. Science* **373**, 1004–1012 (2021).

15. Woo, J. *et al. Science* **376**, 527–532 (2022).

16. Reymond, J.-L. *Acc. Chem. Res.* **48**, 722–730 (2015).

17. Brown, D. G. & Boström, J. *J. Med. Chem.* **59**, 4443–4458 (2016).

18. Ditzler, R. A. J. & Zhukhovitskiy, A. V. *J. Am. Chem. Soc.* **143**, 20326–20331 (2021).

19. Zhong, H. *et al.* Preprint at ChemRxiv https://doi.org/10.26434/chemrxiv-2023-4mw5f (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01735-1

Books & Arts

- **Surgical wonders and dodgy medical ethics: the Hunterian Museum reopens** [31 May 2023]
  • The London museum is a treasure chest of medical specimens both fascinating and ghoulish — now with a renewed focus on questions about how its collections came to be.

- ARTS REVIEW
- 31 May 2023

Surgical wonders and dodgy medical ethics: the Hunterian Museum reopens

The London museum is a treasure chest of medical specimens both fascinating and ghoulish — now with a renewed focus on questions about how its collections came to be.

- Nisha Gaind



This crocodile hatchling is one of more than 2,000 specimens in the Hunterian Museum.Credit: Hunterian Museum

The Hunterian Museum 38–43 Lincolns Inn Fields, London, UK
hunterianmuseum.org

Eighteenth century Scottish surgeon and anatomist John Hunter made trailblazing contributions to both fields, including performing one of the first

instances of artificially inseminating a woman. But he is perhaps best known today as an inveterate collector who acquired, dissected and preserved human and animal tissues from all corners of Earth.

All in all, Hunter, who lived from 1728 to 1793, filled more than 14,000 specimen jars. Most have been lost over the years, including more than 10,000 destroyed by the bombing of London during the Second World War. But many of those that remain tell poignant stories, from a tiny crocodile hatchling emerging from its egg to a man's bullet-perforated anatomy. Together with objects from Hunter's London homes and items such as surgical saws, hammers and retractors, roughly 2,000 of these surviving specimens form the centrepiece of the Hunterian Museum in London, which reopened in its home at the Royal College of Surgeons of England on 16 May after a six-year, £4.6-million (US$5.8-million) refurbishment.

The revamped museum traces the history of surgery from its barbaric origins in the ancient world to the slick, robot-assisted practices followed in many parts of the world today. For Dawn Kemp, the director of museums who led the refurbishment, its unique offering lies in displaying Hunter's work as a comparative anatomist. He studied forms from across the animal and plant kingdoms — often exotic species that were brought to him from military or exploratory voyages — which lie alongside human organs, often in strange juxtaposition. "He's linking all of the living things on Earth. So it's putting humans into nature, not apart from it," says Kemp.



[Imperialism’s long shadow: the UK universities grappling with a colonial past]

At the heart of the museum is the Long Gallery — an atmospherically lit corridor lined with cabinets containing ethanol-preserved remains glowing ghostly white. They are organized into two groups. There are the healthy specimens, including the brains of a tiger (surprisingly small) and a horse, and a Surinam toad caught in the middle of reproduction. Then there are the unhealthy: organs ravaged by cancer and a collection of grisly specimens that have venereal diseases, one of Hunter’s specialist subjects. Off this corridor are rooms that explore different aspects of Hunter’s career, including his work with his older brother William, who was also an anatomist and physician, and his home life with his wife Anne, a poet.

Hunter clearly had a voracious appetite for knowledge. One exhibit chronicles Hunter’s run-in with the scientific integrity issues of his day: it shows the cancer-stricken breast of a woman who had been treated with a spurious chemical remedy for cancer, called Plunkett’s powder. Hunter was sceptical of such ‘cures’ and confronted its promoter about what proof he had for his claims. “Evidence-based science of any sort is absolutely Hunter’s maxim,” says Kemp.

But drastically different ethics now underpin Hunter’s artefacts, how they were collected and their display — uncomfortable facts that the reimagined museum explicitly acknowledges. For example, one wall features portraits of people of different races who at the time had gained celebrity status as human curiosities. These portaits were originally displayed alongside human and animal specimens in Hunter’s home in London’s Leicester Square. There is some speculation that he invited some of those portrayed to his house, and Hunter’s work explored human variation, which influenced later racial theories. Kemp acknowledges that the Hunterian was a place “where some of those closely involved in the Western ‘colonial project’ developed sinister and awful ideas on racial theory”.



As part of its refurbishment, the museum aims to address the sinister side of Hunter’s legacy.Credit: Hunterian Museum

The gallery’s introductory panel notes that Hunter gathered his specimens before modern standards of consent were established. The concept of an ethical review board was as alien to Hunter as ‘body-snatching’, the gruesome practice through which he procured many remains, is to viewers today. “We recognise the debt owed to those people — named and unnamed — who in life and death have helped to advance medical knowledge,” it reads.

Rather than choosing not to display items whose provenance is uncertain, Kemp’s team has tried to address this legacy by making the effort to research and show the names of previously anonymous people. “We had to have these people represented to be able to understand and tell their story,” she says. The museum has, however, removed from display the 2.3-metre skeleton of the ‘Irish Giant’ Charles Byrne, whose remains Hunter acquired at his death in 1783 for a rumoured £500, despite Byrne’s wish to be buried at sea.

Recognizing that issues of ethics can’t be addressed solely by labels, the museum plans this year to launch Hunterian Provocations, a programme that

will include events, workshops and exhibitions that explore themes such as decolonization, the ethics of display, and inclusion and diversity in surgery. "We know these things were wrong. How do we learn and repair, and have a place for open debate where we really value hearing these perspectives?" says Kemp.



Alfred Russel Wallace's first expedition ended in flames

Another of Kemp's guiding principles was to emphasize surgery as a human story and to feature patients' experiences. This shines in the museum's final galleries, which explore modern surgery in brightly lit rooms reminiscent of an operating theatre. Video interviews tell the stories of 11 people who have undergone life-changing surgeries — including a mother whose baby with spina bifida was operated on *in utero* (the videos are viewable on the museum's website).

Alongside is an arresting final specimen: the excised heart of Jennifer Sutton, a woman with a congenital heart condition who received a transplant 16 years ago in her early 20s. Sutton leapt at the chance to have her original heart exhibited. "Surgeons and doctors have innovated and been able to get to a point where I can have a heart transplant," she says. "It allows people to hopefully get to the end of the exhibition and think, look at everything that's come before me. I wouldn't be here if it wasn't for that."

Sutton’s condition is one that might, in the not-too-distant future, be treated by gene-editing techniques, negating the need for an operation. But the universality of surgery, and the opportunity to view our common, bloody building blocks, is what Kemp hopes will draw people to the Hunterian. “You’re looking at human body parts that are the same as you have,” she says. This visceral museum is “a unique place to contemplate what it is to be human”.

Nature **618**, 25-26 (2023)

doi: https://doi.org/10.1038/d41586-023-01736-0

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01736-0

| Section menu | Main menu |

Opinion

- **Why the world needs more transparency on the origins of novel pathogens** [30 May 2023]
  • Collaboration and openness are essential to minimize the risks of future pandemics, says the World Health Organization's scientific advisory group, SAGO.

- COMMENT
- 30 May 2023

Why the world needs more transparency on the origins of novel pathogens

Collaboration and openness are essential to minimize the risks of future pandemics, says the World Health Organization's scientific advisory group, SAGO.

- Marietjie Venter



Researchers catch bats emerging from a cave in Thailand to test them for coronaviruses. Credit: Adam Dean/New York Times/Redux/eyevine

On 5 May, the director-general of the World Health Organization (WHO), Tedros Adhanom Ghebreyesus, declared that COVID-19 is no longer a public-health emergency of international concern. This signalled the end of the acute phase of a pandemic that has caused almost 7 million confirmed deaths worldwide. But more than three years after COVID-19 was declared a pandemic, where the coronavirus SARS-CoV-2 came from, and how it first came to infect humans, remains unclear.

Understanding the origins of novel pathogens that could result in major outbreaks, epidemics or pandemics is essential to prepare for the next emerging disease. In recognition of this, the WHO established the Scientific Advisory Group for the Origins of Novel Pathogens (SAGO) in 2021. The 27 members of our group, from 27 countries, include specialists in epidemiology, virology, human and animal infectious diseases, ecology, genomics, and biosafety and biosecurity.



WHO declares end to COVID-19's emergency phase

Since SAGO was established, we have evaluated what new data are available to aid investigation of the origin of SARS-CoV-2, met with scientists from around the world and made recommendations about the kinds

of data and studies that still need to be collected or pursued. Another key task has been to develop the WHO Global Framework to define and guide studies into the origins of any emerging or re-emerging pathogens of epidemic and pandemic potential. This document is currently under review and due to be published later this year. It is the first attempt to bring together a broad suite of experiences from investigations into the origins of multiple high-risk pathogens, including SARS-CoV-2, mpox (an infectious disease caused by the monkeypox virus), novel influenza viruses, Marburg and Ebola.

The effectiveness of our ongoing efforts to guide investigations into the origins of high-risk pathogens, however, depends on robust science following key principles of outbreak investigation, as well as collaboration, transparency and trust among all stakeholders. To minimize the risk of further disruptive outbreaks, and their associated mortality and burdens on health and economics, various groups must come together — including scientists, public-health officials, governments and global health agencies.

Data drive

In March 2021, together with Chinese researchers, an international team of experts convened by the WHO produced an initial joint report evaluating four possible ways in which SARS-CoV-2 might have emerged[1]. Since November 2021, our advisory group has met numerous times and assessed what studies and data relevant to the origin of SARS-CoV-2 have become available since that 2021 report, and to determine what other studies are needed.

We have assessed investigations of human cases of respiratory illness in China in the months leading up to the first recognized transmission of SARS-CoV-2 in December 2019. We have explored studies of animals that are susceptible to SARS-CoV-2 or closely related viruses, and cases of reverse zoonosis, in which the virus has jumped from humans to animals. We have examined environmental studies — involving samples and swabs taken from the Huanan Seafood Market in Wuhan, China, which was the early epicentre of the COVID-19 pandemic, and wastewater samples collected later from all over the world. Lastly, we have reviewed

investigations of the genomics and evolution of the virus, as well as biosafety and biosecurity considerations.

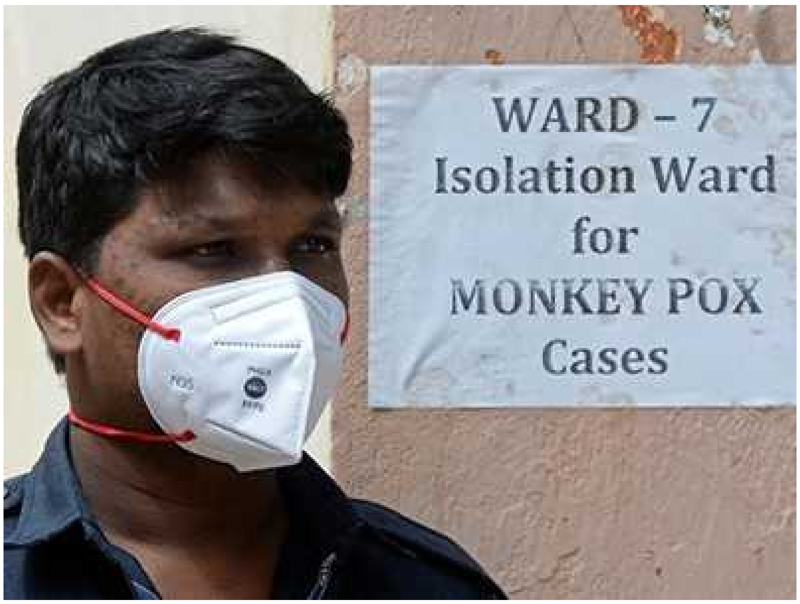


A stall selling food products at a market in Wuhan, China, in January 2021.Credit: Hector Retamal/AFP/Getty

Our ongoing review of the available evidence has made it clear that more work is needed. In June 2022, we published our preliminary report[2], which set out what studies still need to be conducted to understand how the COVID-19 pandemic began. In it, we urged China and other nations to investigate a range of samples that had been collected in the months leading up to the first known cases in December 2019. These include samples taken in research or clinical settings, and as part of environmental surveillance efforts, such as wastewater samples.

We encouraged governments to analyse data from surveillance samples for influenza and severe acute respiratory infections, as well as serology data (from blood banks, for instance) collected during the months before December 2019. The presence of specific antibodies in blood at this time could indicate a previous infection with SARS-CoV-2, which would, in turn,

signal that there had been unnoticed circulation of this new pathogen. We also urged China to investigate animal samples taken as far back as 2003 (when SARS-CoV, the coronavirus that causes severe acute respiratory syndrome, was first detected) from species that can be infected with SARS-like coronaviruses, including bats from across southeast Asia.



'The disease will be neglected': scientists react to the WHO ending mpox emergency

We requested that researchers around the world make any sequencing data from SARS-CoV-2-like viruses available on global genomic databases. We strongly recommended that researchers in China investigate the upstream sources of the animals and animal products that were present at the Huanan Seafood Market before it was closed on 1 January 2020. We also requested access to data or documentation on any possible breaches in biocontainment, or on occupational hazards among lab workers or field investigators from China and other countries working on SARS-like viruses. We have not yet received any new data.

Besides the work on SARS-CoV-2 origins, SAGO has been working with the WHO to develop a global framework to define and guide studies into the origins of emerging and re-emerging pathogens of epidemic and pandemic potential. This framework lays out six areas of work that should be pursued

were any pathogen with epidemic or pandemic potential to emerge or re-emerge in a region where it has not been found before.

These work areas are:

1. Early investigations to establish what activities and factors could have resulted in the initial infections, and to identify how the disease first appeared in humans. These initial investigations should be followed, not necessarily in any particular order, by the following.

2. Investigations that seek to understand the epidemiology, including how the disease presents clinically in humans, and how infections spread.

3. Studies that investigate potential exposures at the animal–human interface. These would establish which animal species are susceptible and determine whether, where and how the virus might have jumped between those species and humans, or from humans to animals.

4. Studies that identify environmental and ecological influences to determine whether the pathogen is being transmitted by vectors, such as mosquitoes or ticks, or through food or water.

5. Genomic and phylogenetic assessments to establish the pathogen's closest ancestors and to track its evolution.

6. Investigations that determine whether lapses in biosafety or biosecurity led to an accidental or intentional release of a pathogen into the human population.

This framework has already been applied to the re-emergence of mpox, as well as to the investigation of the origins of SARS-CoV-2. In December 2022, we published recommendations for investigating the origins of mpox (see go.nature.com/3mnmtpp).

Animals in the mix

In March this year, we were made aware of new sequencing data released by the Chinese Center for Disease Control and Prevention[3], as well as an

independent analysis of metagenomics data by international scientists[4]. The analysis assessed wastewater samples, as well as swabs and animal samples that had been collected at the Huanan Seafood Market in December 2019.

The data suggest that a range of animal species, besides those listed in the March 2021 report[1], were present at the market. Although the data do not identify an animal species that was responsible for the first human case of COVID-19, they do point to various animal species as potential intermediate hosts.



Government workers in Côte d'Ivoire assess chicks in a bid to contain an outbreak of H5N1 avian influenza in 2015.Credit: Luc Gnago/Reuters

In our statement this March, SAGO requested that all sequencing data and any preprints that are in review be shared as soon as possible to allow analysis by the scientific community (see go.nature.com/3mddptv). We also requested that researchers in the United States and elsewhere who are using these data collaborate with Chinese researchers.

We will continue to monitor emerging evidence on the origins of SARS-CoV-2 and provide guidance to the WHO on any new data. Continuing from our June 2022 report[2], we will publish a more comprehensive assessment of current data on the origins of SARS-CoV-2 in the coming months. We will follow up on the recommendations we outlined in June, and will continue to apply the global framework to the investigation of novel pathogens.

For all the work we are doing, however, it is essential that investigations into the origins of novel high-risk pathogens be grounded in science and international collaboration. Such investigations must not be seen as the responsibility of one country, and geopolitics must not be allowed to interfere with or block these.

Throughout the COVID-19 pandemic, attempts to get data out quickly led to the rapid release of preprints, some of which have since been retracted. This has resulted in key stakeholders and the public mistrusting the system. In our view, editors, reviewers and publishers should explore ways of facilitating faster review, to allow rapid sharing of data without erosion of scientific quality.

Platforms such as GISAID (the Global Initiative on Sharing All Influenza Data) have enabled important sequencing data to be shared by scientists, with the aim of also protecting researchers' intellectual property. However, users of these data need to work more closely with the data owners to maintain trust and ensure that sharing continues in the future. Most importantly, we urge scientists and governments to make available all data, research and reports that can help in the identification of the origins of novel pathogens for all outbreaks, epidemics and global health emergencies.

It is only through these means that investigations can be based on scientific rigour — not on speculations that lead to mistrust.

Nature **618**, 27-29 (2023)

doi: https://doi.org/10.1038/d41586-023-01722-6

References

1. World Health Organization. *WHO-Convened Global Study of Origins of SARS-CoV-2: China Part. Joint WHO–China Study* (WHO, 2021).

2. WHO Scientific Advisory Group on the Origins of Novel Pathogens. *Preliminary Report* (WHO, 2022).

3. Liu, W. J. *et al. Nature* https://doi.org/10.1038/s41586-023-06043-2 (2023).

4. Crits-Christoph, A. *et al.* Preprint at Zenodo https://doi.org/10.5281/zenodo.7754299 (2023).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01722-6

Work

- **To diversify the scientific workforce, postdoc recruitment needs a rethink** [26 May 2023]
  • Biased hiring practices are limiting efforts to attract and retain researchers of colour.
- **I use a robot to restore Pompeii’s shattered frescoes** [30 May 2023]
  • Cultural and heritage researcher Arianna Traviglia uses robotic technology to painstakingly reconstruct 2,000-year-old mural paintings in Pompeii.

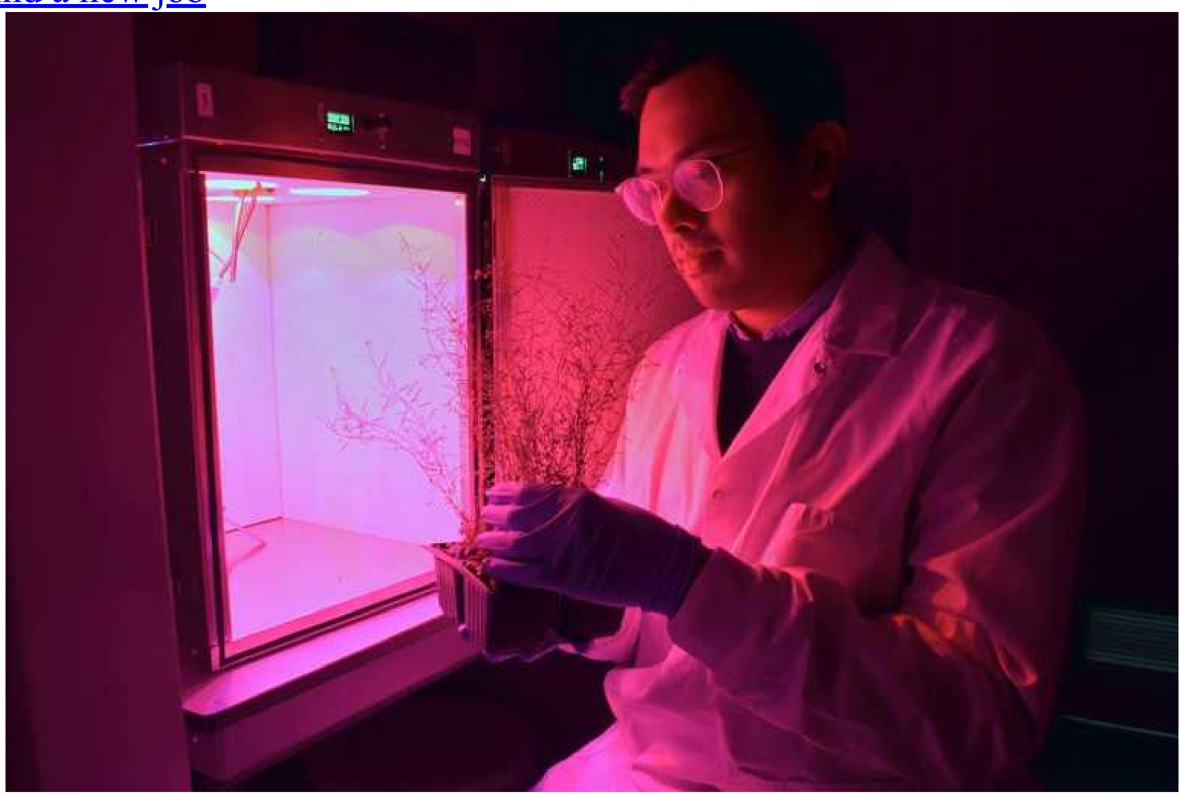
- CAREER FEATURE
- 26 May 2023

To diversify the scientific workforce, postdoc recruitment needs a rethink

Biased hiring practices are limiting efforts to attract and retain researchers of colour.

- Esther Landhuis

Find a new job



Devang Mehta was hired through an ‘open’ faculty-recruitment programme.Credit: Kaat Hebbelinck

It’s been known for nearly half a century that the ‘chilly climate’ of science higher education can make some students of colour feel unwelcome[1]. Since then, the shortage of people from minority groups in science, technology, engineering, and mathematics (STEM) fields has captured public attention and research interest. Diversity initiatives seek to identify and equip under-represented students, and there is growing recognition that the academic environment itself needs reform.

Figures suggest there has been some progress. Between 2006 and 2016, the number of African Americans and Latinos who obtained doctoral degrees in the United States rose by 31% and 54% respectively, according to data from the 2018 US National Science Foundation (NSF) Survey of Earned Doctorates.

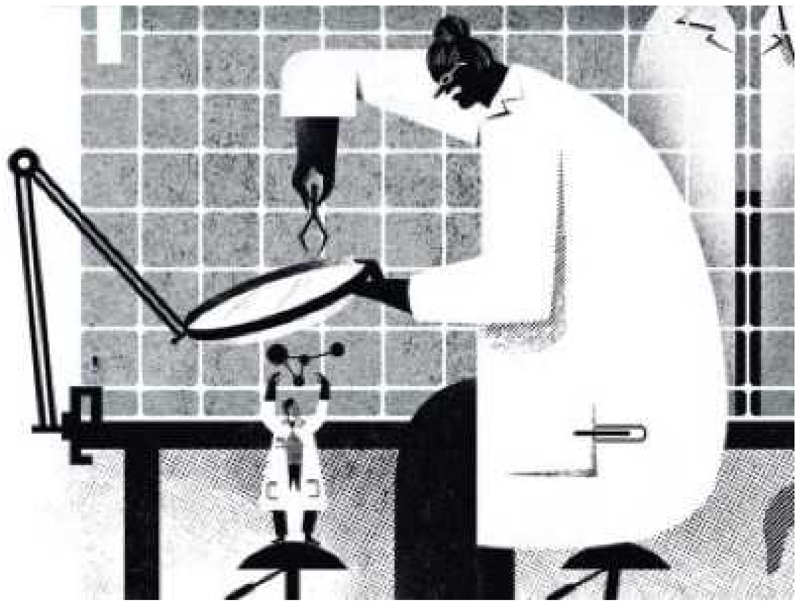
Yet, despite these gains, decades of focus and millions of research dollars have failed to produce meaningful change at the highest levels of academic achievement and influence — the professoriate. In the United States, the share of tenure-track or tenured faculty members from under-represented groups — Black, Latino, American Indian and Alaska Native — edged up just one percentage point (from 11% to 12%) between 2013 and 2019[2]. In the United Kingdom, Black people make up 8% of the country’s science undergraduates but only 0.6% of science professors, according to a December 2022 *Nature* investigation (see go.nature.com/3nk3c5h).

Since 2020, both the global protests by the Black Lives Matter movement and time for reflection during COVID-19 pandemic lockdowns have helped to reprioritize and reinvigorate long-standing efforts to diversify science.

“People realized that this is the time to create a change,” recalls Mayank Chugh, who arrived in Boston, Massachusetts, to begin a systems biology postdoc at Harvard Medical School several months before the pandemic hit.

Researchers who analysed the biomedicine training pathway found that the greatest loss of people from under-represented minorities seems to occur at two points: during undergraduate education, and in the transition from

postdoctoral fellowship to tenure-track faculty position[3]. The latter juncture is particularly challenging: postdocs work long hours for little pay, and those from marginalized backgrounds face extra hurdles because of undue bias in faculty hiring.



Career resources for postdoctoral researchers

Nature's 2020 global survey of postdoctoral researchers revealed that more than half (53%) of the 7,670 respondents reported earning less than US$50,000 annually, and 61% typically worked at least 6 hours a week more than their contracted hours (see go.nature.com/45kjk2v). (Look out for the chance to take *Nature*'s 2023 global postdoc survey later this month.)

Efforts to alleviate these barriers focus on creating clearer processes for admission to postdoctoral positions, increasing the visibility of postdocs' research, and involving a broader range of people on faculty hiring committees, including having applicants' research plans evaluated by experts who are not part of the hiring department.

Current practices for STEM faculty recruitment are "designed around white men", says Marybeth Gasman, education researcher at Rutgers University in New Brunswick, New Jersey, and author of a book on tackling racism in academic hiring, *Doing the Right Thing* (2022). "We have to figure out how to redesign [them]."

No standard path

Unlike undergraduate or graduate programmes, which have a structured admissions process, many postdoc positions are filled by an informal process, in some cases by word of mouth. A faculty member who is awarded a grant to hire a postdoc will typically say, “I need someone with this expertise, so I’m going to share that with my network and see if they have emerging students who might be interested,” says Jessica Bennett, assistant vice-president of STEM education at the Association of Public and Land-grant Universities in Washington DC.

Postdoctoral candidates will often express interest in an advertisement or post by sending an e-mail to the principal investigator (PI). It proceeds informally from there, says Chugh. If it seems like a good match and an initial chat goes well, the PI will invite the candidate to give a research talk and meet laboratory members before deciding whether to offer them the job.



The postdoc career stage can be key to career advancement, says Jessica Bennett.Credit: Josh Gold Photography

Given the lack of oversight and consistency for this crucial step of the pipeline, it's hard to nail down trends in the hiring of minority postdoctoral candidates. "Who is applying? Where are they from? What are their stories like?" asks Chugh. This would help to inform processes and decision-making, but "we don't have the data", he says.

Those data would mainly come from individual PIs, says Chugh. "We want to see, for example, if a PI gets 100 applications, who are these applications from? Are there biases around who gets an interview? We just don't know. Everything is at the discretion of the PI and their lab."

Other complications arise from inconsistencies in how postdocs are categorized. Some institutions treat postdocs as lecturers or academic staff, whereas other places might consider them as more like a "student on steroids", Bennett says. That makes it challenging to address the loss of under-represented minority trainees during the postdoc stage.

High stakes

Time spent as a postdoc is a crucial step for career advancement. It's when a trainee transitions from working on problems outlined by the PI to thinking more independently — "developing new lines of research, establishing themselves in their field", says Bennett. The stakes are high: in most STEM fields, the postdoc stage "is considered almost a requirement for hiring into the faculty position", she says.

For institutions, postdocs serve another purpose — "quite frankly, getting inexpensive labour to work in labs", says Gasman, who also serves as Rutgers' associate dean for research in the Graduate School of Education.

Gasman also spent more than a decade co-directing a postdoc programme, funded by the US National Institutes of Health (NIH), that focused on giving postdocs teaching experience at institutions that serve large numbers of students from minority ethnic groups. She found that postdocs would, at times, come to her and "confide that they were being worked countless hours", she says. Many were spending 80 or more hours a week doing lab research, while also mentoring PhD students and undergraduates, leaving

few hours for family if they were married and had a child, Gasman says. It's easy for postdocs to feel trapped.



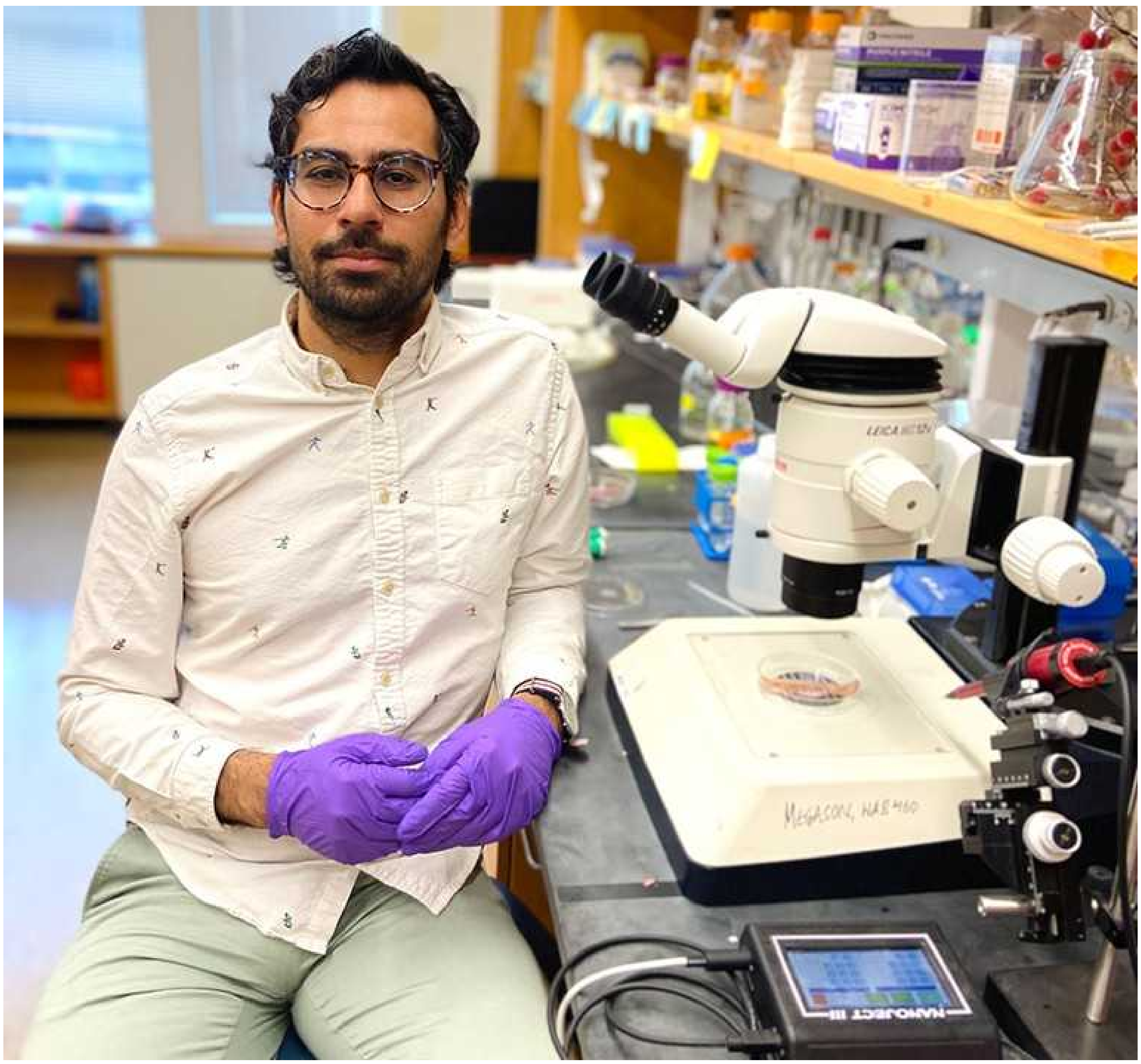
A postdoc's guide to choosing the right lab

"They're really at the mercy of the PI, because if they don't do a good job, they can't get a faculty position," Gasman says.

Another challenge is low pay for the heavy workload. Although annual salaries for entry-level US postdocs have steadily risen over several decades to $54,840, according to 2022 NIH guidelines, this is still less than the median income ($58,396) for full-time workers spanning all levels of education. Across the globe, the low pay — combined with factors such as demanding hours and uncertain career prospects — has steered more doctoral graduates away from academic postdoc positions.

Financial considerations probably do stymie efforts to diversify the STEM workforce, says Kimberly Griffin, dean of the College of Education at the University of Maryland in College Park. Students who are finishing their doctoral degrees and looking to a postdoc role are thinking to themselves, "I have years and years of education, and I'm going into a position that's not going to pay me the same as if I was working in another field", she says. "When we look at who holds the most debt, it's often African American students, and Black women in particular," she adds.

And yet, finances were not the prevailing factor in an analysis undertaken by Griffin and Kenneth Gibbs, then with the Directorate for Education and Human Resources at the NSF in Arlington, Virginia, to determine what shapes the career trajectories of biomedical science PhD graduates[4]. For people of colour, alignment of personal values and workplace culture seems to be the key driver, their research found. What could these considerations look like? “The climate isn’t particularly comfortable. I don’t see people who look like me. I’m not sure if people really value the research that I want to do. I’m really distant from the community impact I want to have. And it’s not paying me all that much,” says Griffin. For minority researchers, “these barriers pile up and make someone less likely to pursue that [postdoc]”.



Mayank Chugh has tested a centralized system that collects more-comprehensive data on postdoc applications.Credit: Rachael Jonas-Closs

Although there are a growing number of statements and initiatives aimed at tackling systemic inequities to lower those barriers, it is hard to distinguish what is merely about head counts from what actually creates a sense of belonging, says Chugh, who is president of the Harvard Medical Postdoc Association. Labs and workplaces can diversify "simply by having people with different-coloured skin or having differently abled people", he says. "But are you actually making sure the culture is changing so that the people being hired in these new places do feel welcome?"

Grow your own

It's a tough question. In recent years, some universities have tried to diversify their faculty by recruiting people from minority ethnic groups to 'grow your own' programmes that aim to convert postdocs into faculty members at the same institution[5]. Gasman sees more and more of these programmes cropping up at top research institutions. "All you have to do is search for 'diversity postdoc programme'. You'll find hundreds," she says. But are they effective?



A lab leader's guide to hiring a postdoc

As part of her book research, Gasman interviewed more than 100 people, including provosts, deans and department chairs who were involved in faculty recruiting and hiring at American Association of Universities (AAU) institutions. Many interviewees remained anonymous so that they could speak freely about their institution's hiring practices. Gasman asked whether their university had a diversity postdoc programme and whether it had improved gender and racial parity among the faculty. "A lot of them said, well, 'we started this off as a feeder programme, but we don't end up hiring people'" from it, Gasman says.

Most of these programmes are indeed motivated by considerations of diversity, equity and inclusion, she learnt, but in some cases the hiring committees decided that the postdocs who were recruited weren't 'the right fit' for the department. "They might say it that way," Gasman quips.

The prestige factor

Gasman also asked about the qualities that hiring committees value the most, and found that it boiled down to two things: PhD and postdoc institutional affiliation and the reputation of the candidate's PhD adviser. "People just flat out told me, 'At the end of the day, even if someone doesn't have the publications that we really want, if they have a really well-known adviser that we respect and they went to the right institution, we'll take them,'" Gasman says.

Last year, researchers showed that 80% of tenure-track faculty in the United States came from 20% of PhD-granting institutions[6]. Demographic data from the Survey of Earned Doctorates also suggest that perceived prestige of training influences academic hiring.

African Americans make up 6% of all PhDs produced in the US from 2011 to 2018, while Latinos represented 5% and white people 48%. Yet the picture changes drastically in the professoriate, where African Americans and Latinos made up just 3.4% and 4.3% of tenured and tenure-track AAU faculty staff, respectively, while white people filled 70.4% of these roles.

The drop in African American and Latino representation in the highest faculty ranks does seem to align with the idea that getting a PhD from a non-AAU institution sets up a “bigger hurdle to be able to pursue faculty positions, especially in STEM”, Gasman says. Between 2011 and 2018, about 44% of white people among tenured and tenure-track faculty earned doctorates from AAU institutions, compared with 40% of Latinos and only 25% of African Americans.

“It’s not that there isn’t the production of people of colour in STEM; there is,” Gasman says. “The problem is where they’re getting degrees. It’s very hard to get a major research-university faculty job if you didn’t attend an AAU institution.” People from under-represented groups who don’t land STEM faculty positions typically go into industry or work for the federal government, and some go into secondary-school teaching.



Neil Lewis has highlighted the shortcomings of diversity initiatives.Credit: K.Nakamura, Cornell University & Y. Peterhans, USC Wrigley Institute for Environmental Studies

News stories about high-profile denials of tenure affecting academics from minority ethnic groups could have an impact, too — one that is difficult to measure. “When you see people who are some of the biggest names in their fields get denied, it makes you wonder whether the sacrifices are even worth it,” says Neil Lewis, a behavioural scientist at Cornell University in Ithaca, New York, who wrote about the issue[2] last year in *Nature Human Behaviour*. “If the giants in the field can be denied, what chance do I have? It makes you seriously consider alternatives.”

Small-step solutions

To get a better handle on the postdoctoral pipeline, Chugh and his co-workers are creating a centralized application system, similar to those in place for graduate school applications, to collect data on who is applying for postdocs — which labs they’re from, their publication records and their demographics — and who is accepted. After piloting the system in his department at Harvard, Chugh says it might expand to include different schools at the university and eventually other universities.

It’s also important to support researchers during their postdoc, a period that often lacks cohesion. Whereas graduate students typically arrive on campus in predictable waves according to the academic calendar, postdocs start at different times of year and without a cohort. “It’s one of those weird stages where it’s really hard to have a community,” says Jonny Coates, an immunology postdoc at Queen Mary University of London. Coates started a group on the online-collaboration platform Slack for UK and European Union postdocs to share information on grants and fellowships, learn about careers outside academia and hear fellow members discussing their research.

For postdocs who are starting to search for faculty jobs, there is an approach that some think reduces bias in hiring. Instead of vying for a specific departmental faculty opening, for which professors in that department would have the final say in who gets chosen, postdocs can apply for so-called open positions that pull from various disciplines. Candidates are selected through interviews with a multidisciplinary panel and a third-party peer review of their proposed research plan. “So maybe it’s a little bit more objective, because it’s someone else making the decision, not the department that

you're going to join," says Devang Mehta, a plant systems biologist who was hired through an open faculty-recruitment programme at the Catholic University of Leuven in Belgium in November 2022. Similar programmes exist in the Netherlands and Switzerland.

Working towards systemic change is a long game — and requires thick skin. "If you want to do work related to dismantling the way we do things," Gasman says, "people will not like you. They will ostracize you."

Coates agrees, noting that academia can force change-makers out in ways that aren't obvious. "You just don't get the spot you need to [produce] the data to get a fellowship, for example."

Still, there is reason for optimism. "I think there are enough of us who feel strongly," says Gasman, "across all racial and ethnic groups, that it's so important to diversify the faculty that it will happen."

Nature **618**, 201-203 (2023)

doi: https://doi.org/10.1038/d41586-023-01740-4

References

1. Palid, O., Cashdollar, S., Deangelo, S., Chu, C. & Bates, M. *Int. J. STEM Ed.* **10**, 2 (2023).

2. Lewis, N. A. *Nature Hum. Behav.* **6**, 610 (2022).

3. Meyers, L. C., Brown, A. M., Moneta-Koehler, L. & Chalkley, R. *PLoS ONE* **13**, e0190606 (2018).

4. Gibbs, K. & Griffin, K. A. *CBE Life Sci. Educ.* **12**, 711 (2017).

5. Culpepper, D. *et al. Front. Psychol.* **12**, 733995 (2021).

6. Wapman, K. H., Zhang, S., Clauset, A. & Larremore, D. B. *Nature* **610**, 120–127 (2022).

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01740-4

- WHERE I WORK
- 30 May 2023

I use a robot to restore Pompeii’s shattered frescoes

Cultural and heritage researcher Arianna Traviglia uses robotic technology to painstakingly reconstruct 2,000-year-old mural paintings in Pompeii.

- Stav Dimitropoulos



Arianna Traviglia is a senior researcher at the Centre for Cultural Heritage Technology at the Italian Institute of Technology in Genoa, Italy. Credit: Francesco Rucci and Francesco Marinelli for *Nature*

As principal investigator of the Centre for Cultural Heritage Technology in Genoa, Italy, I promote the development of new technologies to preserve and protect important archaeological findings. Right now, I am in the field in the ancient Roman city of Pompeii, where I am working with a robot to reconstruct Pompeii's shattered frescoes. These are a type of mural painting on wet plaster. The project is called RePAIR.

Pompeii contains thousands of fresco fragments, and it is close to impossible for a human to reconstruct all the irregular shards into big, meaningful paintings. Technology now allows us to reassemble them from myriad pieces: the robot we use is equipped with two soft humanoid arms, high-definition scanners, cameras and 3D digital-recognition software.

In this picture from January, I am in the House of the Painters at Work, part of a complex of buildings called the House of the Chaste Lovers. Both structures remained hidden from the world for about 1,800 years after the ad 79 eruption. I am holding a colorimeter, a device that helps me to determine and specify the original colours in the frescoes.

In RePAIR, I'm in charge of studying the hyperspectral images of the fresco fragments in storage and comparing them with the paintings on the walls of the House of the Painters at Work. By comparing the spectral similarities, we will be able to work out whether the scattered pieces collapsed from the same wall.

I have a PhD in geomatics, and I initially studied history and archaeology. Archaeology helps us to understand where we come from, and technology enriches that chase. I'm glad I'm a translator of two worlds. My mind runs wild when I think about the fascinating titbits of Roman life that we are still missing because Pompeii's frescoes were smashed up.

Nature **618**, 206 (2023)

doi: https://doi.org/10.1038/d41586-023-01741-3

This article was downloaded by **calibre** from https://www.nature.com/articles/d41586-023-01741-3

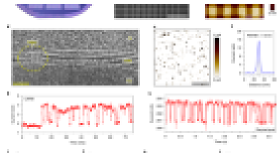
| Section menu | Main menu |

Research

- ## Hybrid 2D–CMOS microchips for memristive applications
  [27 Mar 2023]
  • High-integration-density 2D–CMOS hybrid microchips for memristive applications are made demonstrating in-memory computation and electrical response suitable for the implementation of spiking neural networks representing an advance towards integration of 2D materials in microelectronic products and memristive applications.
- ## Strong and ductile titanium–oxygen–iron alloys by additive manufacturing [31 May 2023]
  • Combining alloy design with additive manufacturing process design creates α–β titanium–oxygen–iron alloys that are both strong and ductile, with the potential to revitalize off-grade sponge titanium and thereby reduce the carbon footprint of the titanium industry.
- ## Enhanced rare-earth separation with a metal-sensitive lanmodulin dimer [31 May 2023]
  • A study biochemically and structurally characterizes a lanmodulin from Hansschlegelia quercus with an oligomeric state sensitive to rare-earth ionic radius.
- ## Ancient gene linkages support ctenophores as sister to other animals [17 May 2023]
  • Deeply conserved syntenic characters unite sponges with bilaterians, cnidarians, and placozoans in a monophyletic clade to the exclusion of the comb jellies (ctenophores)—placing ctenophores as the sister group to all other animals.
- ## Gap junctions desynchronize a neural circuit to stabilize insect flight [24 May 2023]
  • In the Drosophila central-pattern-generating neural network, a mechanism for network desynchronization relying on weak electrical synapses and specific excitability dynamics of the coupled neurons translates unpatterned premotor input into stereotyped neuronal firing with fixed sequences of cell activation, ensuring stable wingbeat power.
- ## Walking naturally after spinal cord injury using a brain–spine interface [24 May 2023]
  • A reliable digital bridge restored communication between the brain and spinal cord and enabled natural walking in a participant with spinal cord injury.
- ## Religion and educational mobility in Africa [17 May 2023]
  • Religion-specific measures of intergenerational mobility in education using census data for 21 African countries indicate that Christians have fared considerably better than their Muslim

or Animist peers even when comparing individuals residing in the same district born to households with similar characteristics.

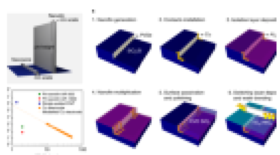
- **Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer** [10 May 2023]
  • A phase I clinical trial of an adjuvant personalized mRNA neoantigen vaccine, autogene cevumeran, in patients with pancreatic ductal carcinoma demonstrates that the vaccine can induce T cell activity that may correlate with delayed recurrence of disease.
- **Uridine-derived ribose fuels glucose-restricted pancreatic cancer** [17 May 2023]
  • A metabolite screen of pancreatic cells shows that pancreatic cancer cells metabolize uridine-derived ribose via UPP1, supporting redox balance, survival and proliferation.
- **Programming inactive RNA-binding small molecules into bioactive degraders** [24 May 2023]
  • Small-molecule RNA-targeted degradation can be leveraged to convert strong, yet inactive, binding interactions into potent and specific modulators of RNA function.



Wafer-scale integration of two-dimensional materials in high-density memristive crossbar arrays for artificial neural networks

19 October 2020

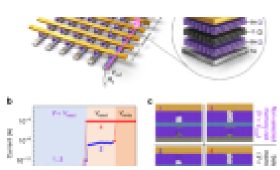
Shaochuan Chen, Mohammad Reza Mahmoodi, … Mario Lanza



Memristor crossbar arrays with 6-nm half-pitch and 2-nm critical dimension

12 November 2018

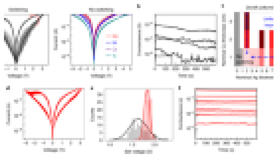
Shuang Pi, Can Li, … Qiangfei Xia



Self-selective van der Waals heterostructures for large scale memory array

18 July 2019

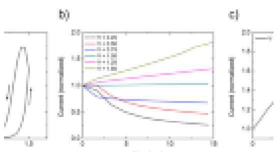
Linfeng Sun, Yishu Zhang, … Heejun Yang



[Alloying conducting channels for reliable neuromorphic computing]

08 June 2020

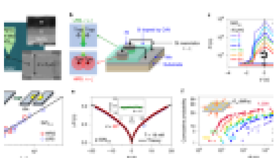
Hanwool Yeon, Peng Lin, … Jeehwan Kim



[Scalable Memdiodes Exhibiting Rectification and Hysteresis for Neuromorphic Computing]

28 August 2018

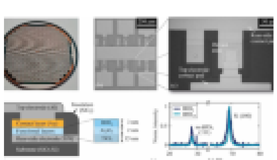
Joshua C. Shank, M. Brooks Tellekamp, … W. Alan Doolittle



[An electronic silicon-based memristor with a high switching uniformity]

15 February 2019

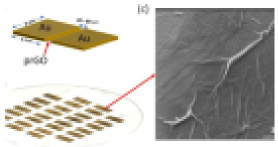
Yang Lu, Ana Alvarez, … I-Wei Chen



[Multilayer redox-based HfOx/Al2O3/TiO2 memristive structures for neuromorphic computing]

29 October 2022

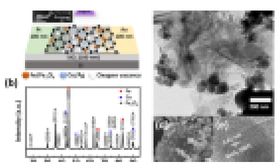
Seongae Park, Benjamin Spetzler, … Martin Ziegler



[NeuroMem: Analog Graphene-Based Resistive Memory for Artificial Neural Networks]

11 June 2020

Heba Abunahla, Yasmin Halawani, … Baker Mohammad



[Highly light-tunable memristors in solution-processed 2D materials/metal composites]

05 November 2022

Zahra Sheykhifar & Seyed Majid Mohseni

- Article
- Open Access
- Published: 27 March 2023

Hybrid 2D–CMOS microchips for memristive applications

- Kaichen Zhu,
- Sebastian Pazos,
- Fernando Aguirre,
- Yaqing Shen,
- Yue Yuan,
- Wenwen Zheng,
- Osamah Alharbi,
- Marco A. Villena,
- Bin Fang,
- Xinyi Li,
- Alessandro Milozzi,
- Matteo Farronato,
- Miguel Muñoz-Rojo,
- Tao Wang,
- Ren Li,
- Hossein Fariborzi,
- Juan B. Roldan,
- Guenther Benstetter,
- Xixiang Zhang,
- Husam N. Alshareef,
- Tibor Grasser,
- Huaqiang Wu,
- Daniele Ielmini &
- …
- Mario Lanza

Show authors

Nature **volume 618**, pages 57–62 (2023)

- 10k Accesses
- 5 Citations

- 108 Altmetric
- Metrics details

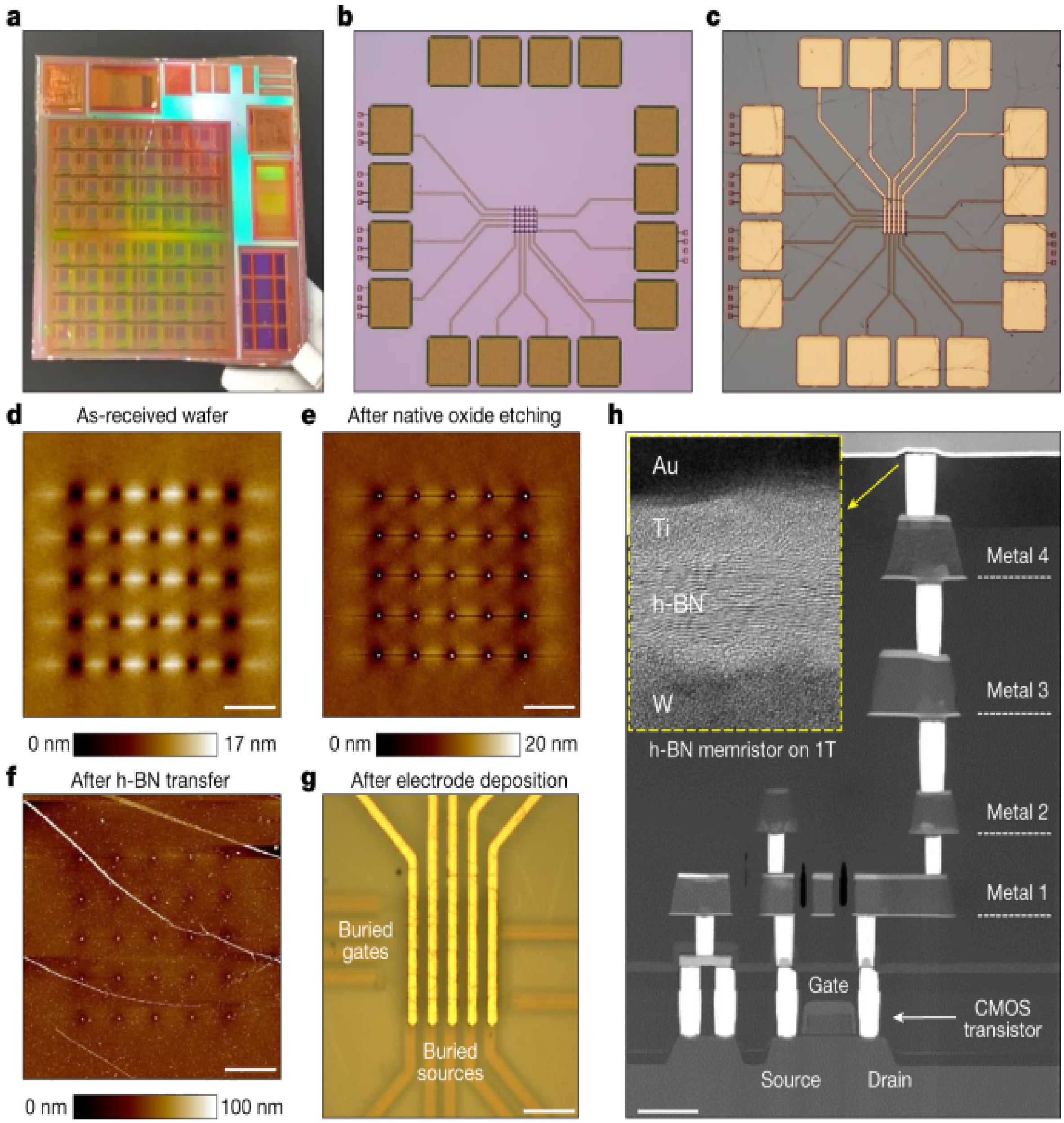
Abstract

Exploiting the excellent electronic properties of two-dimensional (2D) materials to fabricate advanced electronic circuits is a major goal for the semiconductor industry[1,2]. However, most studies in this field have been limited to the fabrication and characterization of isolated large (more than 1 μm^2) devices on unfunctional SiO_2–Si substrates. Some studies have integrated monolayer graphene on silicon microchips as a large-area (more than 500 μm^2) interconnection[3] and as a channel of large transistors (roughly 16.5 μm^2) (refs. [4,5]), but in all cases the integration density was low, no computation was demonstrated and manipulating monolayer 2D materials was challenging because native pinholes and cracks during transfer increase variability and reduce yield. Here, we present the fabrication of high-integration-density 2D–CMOS hybrid microchips for memristive applications—CMOS stands for complementary metal–oxide–semiconductor. We transfer a sheet of multilayer hexagonal boron nitride onto the back-end-of-line interconnections of silicon microchips containing CMOS transistors of the 180 nm node, and finalize the circuits by patterning the top electrodes and interconnections. The CMOS transistors provide outstanding control over the currents across the hexagonal boron nitride memristors, which allows us to achieve endurances of roughly 5 million cycles in memristors as small as 0.053 μm^2. We demonstrate in-memory computation by constructing logic gates, and measure spike-timing dependent plasticity signals that are suitable for the implementation of spiking neural networks. The high performance and the relatively-high technology readiness level achieved represent a notable advance towards the integration of 2D materials in microelectronic products and memristive applications.

Main

Our 2 cm × 2 cm silicon microchips have been designed by means of Synopsys software and fabricated in a 200 mm silicon wafer in an industrial clean room using a 180 nm CMOS technology node (Fig. 1a and Extended Data Fig. 1). The circuits fabricated in this study consist of 5 × 5 crossbar arrays of one-transistor-one-memristor cells (1T1M, Fig. 1b,c and Supplementary Fig. 1), although some standalone memristors and CMOS transistors were fabricated for reference (Supplementary Fig. 2). The microchips have been designed to integrate the memristors into the back-end-of-line (BEOL) interconnections; that is, they have been terminated at the latest metallization layer (fourth in our wafer) and have been left without passivation. Hence, silicon oxide naturally grows on the wafers when they are extracted from the industrial clean room (Fig. 1d), which can be easily etched away to expose the tungsten vias (Fig. 1e and Supplementary Fig. 3). Then, a roughly 18-layer-thick sheet of hexagonal boron nitride (h-BN) (that is, roughly 6 nm), grown on a Cu substrate by means of chemical vapour deposition (CVD), was transferred on the microchips (Fig. 1f) using a low-temperature process (Methods). Finally, the h-BN on the contact pads was etched, and top electrodes made of different materials (that is, Au–Ti, Au or Ag) were patterned and deposited on the h-BN to finalize the circuits (Fig. 1g).

Fig. 1: Fabrication of hybrid 2D–CMOS memristive microchips.



a, Photograph of the 2 cm × 2 cm microchips containing the CMOS circuitry. **b**,**c**, Optical microscope images of a part of the microchip containing a 5 × 5 crossbar array of 1T1 M cells, as received (**b**) and after fabrication (**c**). The size of the squared pads is 50 µm × 50 µm. **d**–**f**, Topographic maps collected with atomic force microscopy of the vias in the 5 × 5 crossbar arrays on the wafers as received (**d**), after native-oxide etching (**e**) and after the transfer of the h-BN sheet (**f**). **g**, Optical microscope image of a finished 5 × 5 crossbar array of 1T1M, that is, after h-BN transfer and top electrodes deposition. **h**, High-angle annular dark-field cross-sectional scanning transmission electron

microscope image of a 1T1M cell in the crossbar array. The inset, which is 20 nm × 16 nm, shows a cross-sectional TEM image of the Au–Ti–h-BN–W memristor on the via; the correct layered structure of h-BN can be seen. Scale bars, **d**–**f**, 10 µm; **g**, 25 µm; **h**, 600 nm.

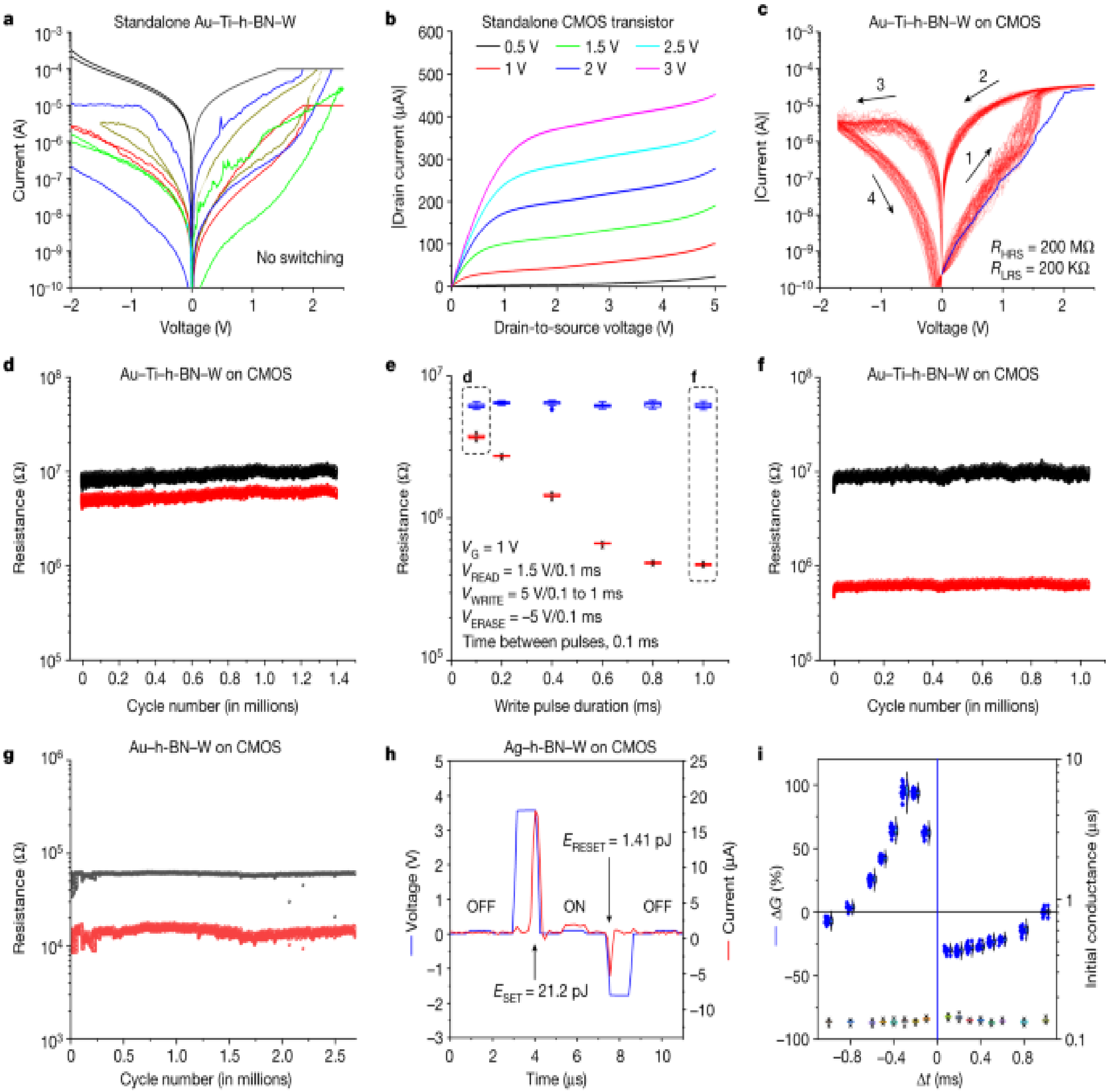
As the tungsten vias of the fourth metallization layer have a diameter of roughly 260 nm (Fig. 1h and Supplementary Fig. 3), the lateral size of the resulting h-BN memristors is, at most, 0.053 μm^2. Figure 1h shows a high-angle annular dark-field cross-sectional scanning transmission electron microscope image of a 1T1M cell (with top Au–Ti electrode) in the crossbar array (Supplementary Figs. 4 and 5). The correct layered structure of the h-BN stack is confirmed before and after transfer by means of cross-sectional transmission electron microscopy (TEM) (inset in Fig. 1h and Extended Data Fig. 2). Nano-chemical analyses by means of electron energy loss spectroscopy demonstrate the correct composition of the h-BN sheet (Extended Data Fig. 3). The optical microscope images (Fig. 1c) reveal that the h-BN sheet does not crack during the transfer; this is an important advantage of using roughly 6-nm-thick 2D layered materials, and it increases the yield of the devices and circuits compared to counterparts using monolayer 2D materials[6].

Electronic memory

When we apply sequences of ramped voltage stresses (RVS) to several standalone 0.053 μm^2 Au–Ti–h-BN–W structures, most (roughly 90%) of them show erratic current fluctuations and no resistive switching (RS) is observed (Fig. 2a); the currents do not reach linear regime (that is, dielectric breakdown) even if we apply 11 V. This is striking because most (more than 75%) h-BN devices with larger areas (25 μm^2) show dielectric breakdown voltages (V_{DB}) between 3 and 11 V followed by filamentary non-volatile bipolar RS[6]. The reason should be the lower probability to find clusters of defects in small devices, which remarkably increases V_{DB} (ref. [7]). Few (roughly 10%) 0.053 μm^2 Au–Ti–h-BN–W structures show V_{DB} between roughly 2.5 and 4 V followed by filamentary non-volatile bipolar RS (if a current limitation 1 mA or higher is applied, Supplementary Fig. 6). However, the endurance is only around 100 cycles, mainly due to the poor

controllability of the current across the memristor and the overshoot during the dielectric breakdown[6].

Fig. 2: Electrical characterization of h-BN–CMOS based 1T1M cells.



a, Electrical characterization of 0.053 μm^2 Au–Ti–h-BN–W structures, showing erratic current fluctuations and no stable RS (each colour line corresponds to one RVS with two polarities). **b**, Typical output characteristic of all standalone CMOS transistors. **c**, Typical non-volatile bipolar RS measured in most 1T1M cells with an Au–Ti–h-BN–W memristor and a CMOS transistor (when applying $V_G = 1.1$ V). **d**–**f**, Endurance plots of 1T1M cell showing around 1.4 million cycles and 1 million cycles for write

pulse durations of 0.1 ms (**d**) and 1 ms (**f**). **e**, R_{LRS} and R_{HRS} in a 1T1M cell when applying pulsed voltage stresses of different durations. **g**, Endurance plot showing non-volatile bipolar RS at $V_G = 1$ V for 1T1M cells using Au–h-BN–W memristors. All the endurance tests have been conducted following the recommended characterization process described in ref. [42]. **h**, Voltage and current versus time in a 1T1M cell with Ag–h-BN–W memristor, showing a low switching energy. **i**, STDP characteristic of the 1T1M cell with Au–Ti–h-BN–W memristor. Before STDP characterization, the devices are always tuned to the same initial conductance (lower box charts, which relate to the right *y* axis).

On the contrary, the CMOS transistor in the 1T1M cell can precisely control the current across the h-BN memristor and avoid the current overshoot, which results in outstanding performance. First, we obtain the output characteristic of one standalone CMOS transistor by applying a constant voltage to the gate (V_G) and a RVS to the drain (V_{DS}), and measuring the drain-to-source current (I_{DS}); the CMOS transistor works correctly as expected (Fig. 2b). And second, we measure the 1T1M cell by applying RVS at the top Au–Ti electrode of the memristor while keeping the source terminal of the transistor grounded and simultaneously applying a constant V_G. When a sequence of RVS is applied to the top electrode of the Au–Ti–h-BN–W structure using $V_G = 1.1$ V, most 1T1M cells show non-volatile bipolar RS (Fig. 2c and Extended Data Fig. 4). The high state resistances (R_{HRS} of roughly 200 MΩ and R_{LRS} of roughly 200 KΩ)—beneficial to reduce power consumption—the non-linearity of the currents in both states and the progressive state transitions indicate that the RS is non-filamentary[8]. However, we do see an activation process, as the first RVS slightly increases the conductance of the devices (that is, softly degrades the h-BN stack, blue line in Fig. 2c and Extended Data Fig. 4). In the first microchip that we fabricated, this stable non-filamentary bipolar RS regime was observed in 32 out of 40 cells (yield 80%), and in the last one it was observed in 25 out of 25 devices (yield 100%). On the contrary, standalone Au–Ti–h-BN–W structures of 0.053 μm^2 and 1T1M cells without h-BN never showed this behaviour; this confirms that the RS is produced by the h-BN stack and that the CMOS transistor is key to control its soft degradation. Note that in the standalone Au–Ti–h-BN–W structures the current is limited using the

semiconductor parameter analyser, which activation time is long (roughly 70 µs) and parasitic capacitance is high (roughly 300 pF, related to the cables)[9]; on the contrary, in the 1T1M cell the series transistor acts as an instantaneous current limitation (it cannot drive more current than that allowed by the size of its channel) and the parasitic capacitance is much lower (roughly 50 pF, internal connections in the microchip), which reduces the duration of the switching transient and undesired currents across the Au–Ti–h-BN–W structure[10]. The values of R_{HRS} and R_{LRS} are stable over time, and multiple stable conductance levels can be programmed either by adjusting V_G during the set process (which fixes R_{LRS}) and/or by adjusting the end voltage of the negative RVS (which fixes R_{HRS}, Extended Data Fig. 5).

The most surprising observation, however, relates to the endurance, which readily reaches 2.5 million cycles (Fig. 2d–f) when applying sequences of pulsed voltage stresses. Under this type of stress, the values of R_{HRS}, R_{LRS} and R_{LRS}/R_{HRS} can be accurately controlled in three different ways: by tuning the duration of the write pulse, by tuning the amplitude of the write pulse, and by tuning the amplitude of the erase pulse (Fig. 2d–f and Extended Data Fig. 6). This endurance is very high considering the small size of the memristors (Supplementary Note 2), and similar to that of commercial metal-oxide-based resistive random access memories (0.5 million cycles)[11] and phase-change memories (10 million cycles)[12,13]. However, the switching time of the 1T1M cells using top Au–Ti electrodes is rather long (t_{SET} of 232 µs and t_{RESET} of 783 ns, Extended Data Fig. 7).

The properties of the 1T1M cells can be adjusted using different top electrodes (Extended Data Fig. 8). When Au electrodes are used, the devices show reliable switching at lower state resistances (Fig. 2g), as well as shorter switching time (t) and lower switching energy (E), and when Ag electrodes are used, these values can be pushed down to t_{SET} = 680 ns, t_{RESET} = 60 ns, E_{SET} = 21.11 pJ and E_{RESET} = 1.41 pJ (Fig. 2h). The reasons behind these observations are the lack of an interfacial Ti layer (which is prone to absorb oxygen, increasing the out-of-plane resistance) and the higher conductivity and diffusivity of Au^{X+} and Ag^{X+} ions[14] (Supplementary Note 1). The performance observed in h-BN–CMOS 1T1M cells using Au–Ti electrodes

may allow them to cover niche applications between NAND Flash and DRAM within the memory hierarchy (for example, persistent memory), and when using Au or Ag electrodes their performance may be valid for low-power application-specific integrated circuits within the internet-of-things[15] (Supplementary Fig. 12).

Data computation

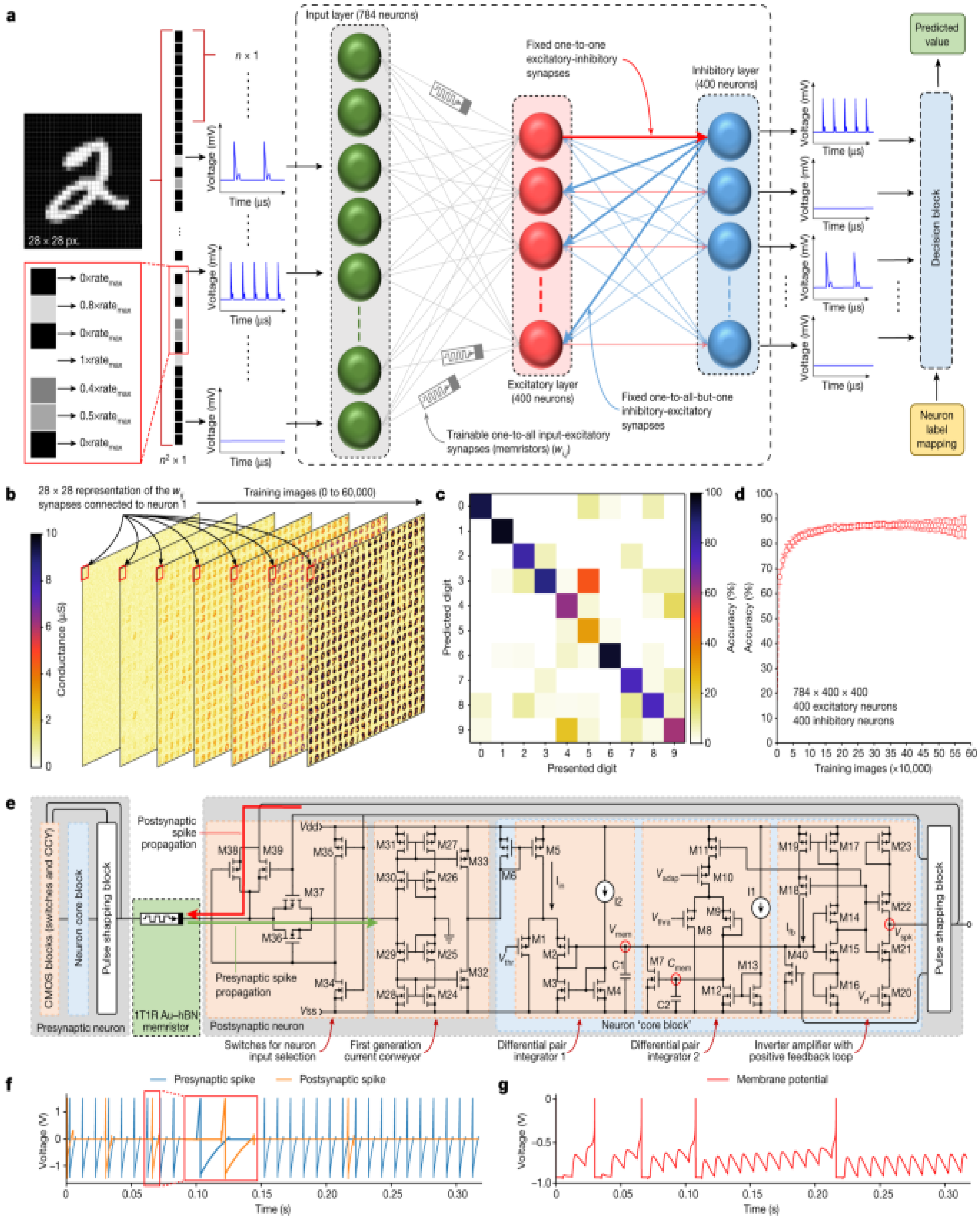
On the basis of the above measured performance metrics, the hybrid 2D–CMOS 1T1M cells show good potential for data computation. The high R_{HRS}/R_{LRS} ratio and the stability of the resistive states over time allows us to implement in-memory computing operations taking advantage of the internal connections of our 5 × 5 crossbar array of 2D–CMOS 1T1M cells. As a proof of concept, we realized 'or' and 'implication' operations (Extended Data Fig. 9), although more sophisticated operations could be easily realized by modifying the interconnections between the devices by means of custom design.

Furthermore, the 1T1M cells with Au–Ti–h-BN–W memristors show spike-timing dependent plasticity (STDP) when applying pairs of pulsed voltage stresses displaced in time at the input and output (Fig. 2i). This non-volatile RS performance is very attractive to construct electronic synapses for spiking neural networks (SNNs)[16], which consume less energy than traditional deep neural networks[17].

Although implementing by means of hardware a reliable 2D materials-based memristive SNN capable of competing with state-of-the-art developments[18,19] is not yet achievable due to the lower maturity of these materials, we can analyse the performance of a SNN made of memristors that show STDP characteristics such as those in Fig. 2i (Supplementary Note 3). First, we fit the measured STDP data from Fig. 2i, including the device-to-device variability, using an exponential decaying model to implement the learning rule (Supplementary Fig. 13). Second, we simulate a SNN to demonstrate the unsupervised learning capability (Fig. 3a), and benchmark it by classifying the images from the Modified National Institute of Standards and Technology (MNIST) database of handwritten digits[20,21] (Methods). The SNN has 784 input neurons, an excitatory layer of 400

neurons and an inhibitory layer of 400 neurons, plus a decision block that determines which is the most probable digit (0–9) represented by the input pattern. We trained the SNN with the complete MNIST dataset and evaluated the accuracy every 1,000 images. Figure 3b–d shows the main three figures-of-merit for this type of SNNs (that is, evolution of the synaptic weights with the number of training images, confusion matrix of the network and the training accuracy versus number of training images) and all of them indicate an excellent performance. To account for the device variability, we considered a Monte Carlo simulation with 50 iterations that randomizes the exponential fitting of the STDP plot and the initial value of the synapses, and the deviations observed in the accuracy are very low (less than 5%, Fig. 3d and Supplementary Fig. 14). The best average accuracy reaches roughly 90%, which is a very high value considering the simplicity of the SNN and the unsupervised training protocol (Supplementary Table 4).

Fig. 3: Implementation of a SNN using CMOS–h-BN based 1T1M cells.



a, Structure of the considered SNN. Each MNIST image is reshaped as a 784 × 1 column vector, and the intensity of the pixels is encoded in terms of the firing frequency of the input neurons. The only trainable synapses are those connecting the input layer with the excitatory layer, and they are modelled with the STDP characteristic of the CMOS–h-BN based 1T1M

cells. The learning is unsupervised, and the neurons are labelled only after the training. These label-neuron assignments are then feed to the decision block altogether with the firing patterns of the neurons, to infer the class of the image presented in the input. **b**, Evolution of the synaptic connections between the input and excitatory layers during training for the case of 400 excitatory and/or inhibitory neurons. The red square identifies 784 synapses arranged in a 28 × 28 representation. **c**, Confusion matrix indicating the classification accuracy for each class from the dataset. **d**, Classification accuracy as a function of the number of presented training images for the neural network comprising 400 excitatory and/or inhibitory neurons. The error bars show the standard deviation for 50 Monte Carlo simulation runs for every accuracy point. **e**, Circuit schematic of the proposed neuron–synapse–neuron block combining h-BN based 1T1M cells and CMOS circuitry. The colours indicate the complete neuron (grey surrounding box), the core block (light-blue box) and the individual building blocks (light-red boxes). CCY, current conveyor. **f**, SPICE simulation of the pre- and postsynaptic signals applied to the CMOS–h-BN based 1T1M. **g**, SPICE simulation of the neuron's membrane potential. The firing events progressively separate from each other due to the adaptative firing threshold.

We also propose a CMOS circuit for the hardware implementation of an electronic neuron based on our h-BN memristors (Fig. 3e), which is capable of accounting for the adaptative firing threshold and the refractory period after firing (pre- and postsynaptic traces and the evolution of the membrane potential, simulated by means of SPICE, in Fig. 3f,g).

Discussion

Very few commercial electronic products today already include 2D materials, and the ones that do (sensors[22], specialty cameras[23]) use very low integration density (more than 100 μm^2 per device)—because in larger devices the local defects in the 2D material are not so detrimental. Our hybrid 2D–CMOS microchips are still far from being ready for production, but we can safely claim that our work represents the highest performance and technology readiness level ever achieved in high-integration-density 2D materials-based electronic devices or circuits. The electrical characteristics of the h-BN memristors connected to a CMOS transistor are by orders of

magnitude superior to those of standalone h-BN memristors[6,24,25,26,27] and h-BN memristors connected to 2D materials-based transistors[28,29].

The voltages needed to switch our devices (from ±1.4 to ±5 V) are low compared to other prototypes in the field of 2D materials (even more than 20 V)[30,31,32], but still higher than that used at the 180 nm CMOS node. Nevertheless, this is not an impediment for the development of this technology, as there are many commercial microchips that operate at much higher voltages; that is the case for all Flash memories[33] (state-of-the-art 3D-NAND Flash memories are programmed at around 20 V)[34] and all bipolar-CMOS microchips for automotive applications (which require up to 40 V)[35]. Strategies to fabricate wafers with devices that operate at different voltages are widespread[36], and many companies[37,38] offer versions of their 180 nm CMOS technology that operate at high voltages greater than 18 V. Note that prototype memristive devices developed by companies also operate at ±5 V (ref. [39]).

We finally remark that, at a first glance, the use of Au and Ag electrodes may not appear ideal because they are categorized as contaminant in front-end-of-line (FEOL) processes. However, our h-BN memristors are integrated in the latest metallic layer of the BEOL interconnections (Fig. 1h), where Au pads, liners and wires are usually used[40] (Extended Data Fig. 10). The semiconductor industry has also developed ferroelectric memories with high content of Iridium[41] (a contaminant material forbidden in FEOL processes), and companies working in the field of 2D materials use Au electrodes in their studies and FEOL prototypes (Supplementary Table 5). Hence, the use of Au, Au–Ti or Ag electrodes in our hybrid 2D–CMOS microchips for memristive applications does not prevent their adoption by the industry.

Methods

Microchip fabrication

The metal-oxide-semiconductor field-effect-transistor circuits were fabricated in a standard CMOS foundry. The size of the wafers is 200 mm,

and the technology node was 180 nm. Each wafer contained 60 chips with a size of 2 cm × 2 cm, and each of them contained different circuits, including the unfinished memristors, finished transistors and unfinished 5 × 5 crossbar arrays of 1T1M cells. As we received the silicon wafer without a passivation layer, first we cut it to separate the microchips and then etched the native oxide by immersing them in a diluted hydrofluoric acid solution (10:1) for 1 min to etch the native oxide (SiO_2). This step was carried out to expose the conductive tungsten vias to make a good electrical contact with 2D materials. Second, we transferred a sheet of multilayer h-BN (previously grown by CVD on a Cu foil) using a wet transfer method. One layer of polymethyl methacrylate (PMMA) with thickness of around 300 nm was spin coated on the cut h-BN. The PMMA–h-BN–Cu sample was deposited on a $FeCl_3$ solution (0.1 g ml^{-1}) to etch the Cu substrate and, once the Cu disappeared, the resulting PMMA–h-BN sample was washed in diluted HCl solution (1 mol l^{-1} for 1 min) and deionized water (for 1 h). The PMMA–h-BN sample was picked up using the native oxide-free CMOS microchip and dried naturally in a dry box. The PMMA was then removed by immersing the sample in acetone for 24 h. Third, we used photolithography (mask aligner from SUSS MicroTec, model MJB4) to expose the h-BN on top of the metallic pad. Then we used a dry etching method with Ar–O_2 plasma (Plasma Cleaner from PVA TePla America Inc., Model IoN 40) to etch the h-BN (300 W for 10 min) and expose the pads. Finally, we used photolithography, electron beam evaporation (Kurt J. Lesker, model PVD75) and a lift-off process (rinse in acetone for 1 min) to pattern and deposit the top electrodes and/or drain electrodes (3 nm Ti with 40 nm Au on top without breaking the vacuum, or 50 nm Au or 50 nm Ag). The process was simple and reproducible, although we believe it could be considerably improved if optimized methods in an industrial clean room are used. Ideally, the h-BN should be grown in large CVD systems (for example, Aixtron[43]) and transferred on the wafers before cutting them in multiple microchips, using methods like laser debonding[44]. Note that even large companies are still using small (7.6 cm or 3-inch) tube furnaces to grow the h-BN for their prototypes[45]. We also confirm that getting the microchips finalized and etch the passivation film before transferring the h-BN works well.

Device characterization

The morphology of the devices was investigated by an optical microscope (DM 4000M, Leica), AFM (Dimension Icon, Bruker) and TEM (Titan Themis, FEI). The thin lamellae for TEM inspection were prepared using a scanning electron microscope provided with focused ion beam (Helios G4 UX, Thermo Fisher Scientific). The electrical characterization was performed by using two probe stations (both M150, Cascade) connected to different semiconductor parameter analysers: a Keithley 4200 and a Keysight B1500A. All the *IV* curves under DC voltages were collected using the Keithley 4200 in the ramped voltage sweep mode, for which three source-measure units are needed for drain, source and gate. Also, all the *IV* curves under pulse mode were collected by Keysight B1500A with two Waveform Generator/Fast Measurement Units connected to the drain and source. An Agilent E3631A DC Power Supply was used to apply constant voltage stress on the transistor gate as gate voltage for pulse measurement. All the endurance plots were collected using the recommended method described in ref. [42].

SNN simulation

The SNN architecture[20] has been developed using Brian2 (ref. [46]), an SNN simulator written in Python. The learning process is based on the empirical measurement of STDP made in the 1T1M cells combining an Au–Ti–h-BN–W memristor on a CMOS transistor. We considered the variability of the network by running a Monte Carlo engine. We benchmarked the accuracy of the SNN during image classification[20] of the MNIST dataset of handwritten digits[21] under an unsupervised learning scheme. We propose a circuit-level model for the neuron–synapse–neuron system, as well as its implementation in SPICE. A detailed description of the SNN and its performance is given in Supplementary Note 3.

Data availability

The data needed to evaluate the conclusions in this work are publicly available online at https://doi.org/10.5281/zenodo.7607096. The datasets that we used for benchmarking are publicly available in ref. [21]. The training methods are provided in ref. [20].

Code availability

The simulator Brian2 used here is publicly available in ref. [46]. The codes used for the simulations described in Methods are publicly available online at https://doi.org/10.5281/zenodo.7607096.

References

1. Akinwande, D. et al. Graphene and two-dimensional materials for silicon technology. *Nature* **573**, 507–518 (2019).

2. *The International Roadmap for Devices and Systems* (IEEE, 2021); https://irds.ieee.org/editions/2021.

3. Goossens, S. et al. Broadband image sensor array based on graphene–CMOS integration. *Nat. Photon.* **11**, 366–371 (2017).

4. Lin, Y. et al. Wafer-scale graphene integrated circuit. *Science* **332**, 1294–1297 (2011).

5. Han, S., Garcia, A. V., Oida, S., Jenkins, K. A. & Haensch, W. Graphene radio frequency receiver integrated circuit. *Nat. Commun.* **5**, 3086 (2014).

6. Shen, Y. et al. Variability and yield in h-BN-based memristive circuits: the role of each type of defect. *Adv. Mater.* **33**, 2103656 (2021).

7. Chen, A. Forming voltage scaling of resistive switching memories. In *Proc. 71st Device Research Conference* 181–182 (IEEE, 2013).

8. Sawa, A. Resistive switching in transition metal oxides. *Mater. Today* **11**, 28–36 (2008).

9. 2600B Source Measure Units (SMU) Instruments. *Tektronix* https://www.tek.com/en/datasheet/2600b-source-measure-units-smu-instruments (2023).

10. Chen, A. Current overshoot during set and reset operations of resistive switching memories. In *Proc. 2012 IEEE International Reliability Physics Symposium (IRPS)* MY.2.1–MY.2.4 (IEEE, 2012)

11. Non-volatile memory with very small operating current ReRAM. *Fujitsu* https://www.fujitsu.com/jp/group/fsm/en/products/reram/ (2023).

12. Analyzing Intel-Micron 3D XPoint: The Next Generation Non-Volatile Memory. *Anandtech* https://www.anandtech.com/show/9470/intel-and-micron-announce-3d-xpoint-nonvolatile-memory-technology-1000x-higher-performance-endurance-than-nand (2023).

13. Kau, D. C. et al. Stackable cross point phase change memory. In *Proc. 2009 IEEE International Electron Devices Meeting (IEDM)* 1–4 (IEEE, 2009).

14. Yang, Y. et al. Electrochemical dynamics of nanoscale metallic inclusions in dielectrics. *Nat. Commun.* **5**, 4232 (2014).

15. Lanza, M. et al. Memristive technologies for data storage, computation, encryption, and radio-frequency communication. *Science* **376**, eabj9979 (2022).

16. Yang, J., Strukov, D. & Stewart, D. Memristive devices for computing. *Nat. Nanotech.* **8**, 13–24 (2013).

17. Merolla, P. A. et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. *Science* **345**, 668–673 (2014).

18. Li, C. et al. Efficient and self-adaptive in-situ learning in multilayer memristor neural networks. *Nat. Commun.* **9**, 2385 (2018).

19. Ishii, M. et al. On-chip trainable 1.4M 6T2R PCM synaptic array with 1.6K stochastic LIF neurons for spiking RBM. In *Proc. 2019 IEEE International Electron Devices Meeting (IEDM)* 14.2.1–14.2.4 (IEEE, 2019).

20. Diehl, P. U. & Cook, M. Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Front. Comput. Neurosc.* **9**, 1662–5188 (2015).

21. LeCun, Y., Cortes, C. & Burges, C. MNIST handwritten digit database (ATT Labs, 2010); http://yann.lecun.com/exdb/mnist/

22. Graphenea card. *Graphenea* https://www.graphenea.com/ (2023).

23. Graphene Flagship industrial partner Emberion launches wide spectrum array sensor at Laser World of Photonics. *Graphene Flagship* https://graphene-flagship.eu/graphene/news/emberion-launches-hyperspectral-array-sensor-at-laser-world-of-photonics/ (2019).

24. Kim, M. et al. Analogue switches made from boron nitride monolayers for application in 5G and terahertz communication systems. *Nat. Electron.* **3**, 479–485 (2020).

25. Shi, Y. et al. Electronic synapses made of layered two-dimensional materials. *Nat. Electron.* **1**, 458–465 (2018).

26. Qian, K. et al. Hexagonal boron nitride thin film for flexible resistive memory applications. *Adv. Funct. Mater.* **26**, 2176–2184 (2016).

27. Zhuang, P. et al. Nonpolar resistive switching of multilayer-hBN-based memories. *Adv. Electron. Mater.* **6**, 1900979 (2020).

28. Yang, R. et al. Ternary content-addressable memory with MoS_2 transistors for massively parallel data search. *Nat. Electron.* **2**, 108–114 (2019).

29. Wang, C. et al. 3D monolithic stacked 1T1R cells using monolayer MoS_2 FET and hBN RRAM fabricated at low (150 °C) temperature. In *Proc. 2018 IEEE International Electron Devices Meeting (IEDM)* 22.5.1-22.5.4 (IEEE, 2018).

30. Wu, F. et al. Vertical MoS_2 transistors with sub-1-nm gate lengths. *Nature* **603**, 259–264 (2022).

31. Liu, L. et al. Uniform nucleation and epitaxy of bilayer molybdenum disulfide on sapphire. *Nature* **605**, 69–75 (2022).

32. Marega, G. et al. Logic-in-memory based on an atomically thin semiconductor. *Nature* **587**, 72–77 (2020).

33. Goda, A. 3-D NAND technology achievements and future scaling perspectives. *IEEE Trans. Electron Devices* **67**, 1373–1381 (2020).

34. Lin, W. L. et al. Grain boundary trap-induced current transient in a 3-D NAND flash cell string. *IEEE Trans. Electron Devices* **66**, 1734–1740 (2019).

35. Automotive Standard Analog Robustness and Performance. Brochure number BR2206AUTOICS. *ST Microelectronics* https://www.st.com/en/automotive-analog-and-power.html#documentation (2023).

36. Niimi, H., Khamankar, R. & Alshareef, H. N. Method for non-thermally nitride gate formation for high voltage devices. US patent 6,566 B2 (2004).

37. Mixed-Signal/CMOS. *Tower Semiconductor* https://towersemi.com/technology/mixed-signal-cmos/ (n.d.).

38. Minixhofer, R. et al. A 120 V 180 nm high voltage CMOS smart power technology for system-on-chip integration. In *Proc. 22nd International Symposium on Power Semiconductor Devices & IC's (ISPSD)* 75–78 (IEEE, 2010).

39. Saitoh, M. et al. HfO_2-based FeFET and FTJ for ferroelectric-memory centric 3D LSI towards low-power and high-density storage and AI applications. In *Proc. 2020 IEEE International Electron Devices Meeting (IEDM)* 375–378 (IEEE 2020).

40. Goodman, P. Current and future uses of gold in electronics. *Gold Bull.* **35**, 21–26 (2002).

41. Summerfelt S. R. et al. Ferroelectric capacitor stack etch cleaning methods. US patent 7,220,600 B2 (2007).

42. Lanza, M. et al. Standards for the characterization of endurance in resistive switching devices. *ACS Nano* **15**, 17214–17231 (2021).

43. Gomollon-Bel, F. Cost effective, large scale graphene with AIXTRON systems. *Graphene Flagship* https://graphene-flagship.eu/graphene/news/cost-effective-large-scale-graphene-with-aixtron-systems/ (2019).

44. Phommahaxay, A. et al. The growing application field of laser debonding: from advanced packaging to future nanoelectronics. In *Proc. 2019 International Wafer Level Packaging Conference (IWLPC)* 1–8 (IEEE, 2019).

45. Chen, T. A. et al. Wafer-scale single-crystal hexagonal boron nitride monolayers on Cu (111). *Nature* **579**, 219–223 (2020).

46. Stimberg, M., Brette, R. & Goodman, D. F. Brian 2, an intuitive and efficient neural simulator. *eLife* **8**, e47314 (2019).

Acknowledgements

This work has been supported by the Ministry of Science and Technology of China (grant nos. 2019YFE0124200 and 2018YFE0100800), the National Natural Science Foundation of China (grant no. 61874075) and the Baseline funding scheme of the King Abdullah University of Science and Technology.

Author information

Author notes

1. These authors contributed equally: Kaichen Zhu, Sebastian Pazos

Authors and Affiliations

1. Materials Science and Engineering Program, Physical Science and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

   Kaichen Zhu, Sebastian Pazos, Fernando Aguirre, Yaqing Shen, Yue Yuan, Wenwen Zheng, Osamah Alharbi, Marco A. Villena, Bin Fang, Xixiang Zhang, Husam N. Alshareef & Mario Lanza

2. Institute of Microelectronics, Tsinghua University, Beijing, China

   Xinyi Li & Huaqiang Wu

3. Department of Electronics, Information and Bioengineering, Politecnico of Milan, Milan, Italy

   Alessandro Milozzi, Matteo Farronato & Daniele Ielmini

4. Department of Thermal and Fluid Engineering, Faculty of Engineering Technology, University of Twente, Enschede, the Netherlands

   Miguel Muñoz-Rojo

5. Institute of Micro and Nanotechnology, IMN-CNM, CSIC (CEI UAM+CSIC), Madrid, Spain

   Miguel Muñoz-Rojo

6. Institute of Functional Nano and Soft Materials, Collaborative Innovation Center of Suzhou Nanoscience and Technology, Soochow University, Suzhou, China

   Tao Wang

7. Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

   Ren Li & Hossein Fariborzi

8. Department of Electronics and Computer Technology, Faculty of Sciences, University of Granada, Granada, Spain

   Juan B. Roldan

9. Department of Electrical Engineering and Media Technology, Deggendorf Institute of Technology, Deggendorf, Germany

   Guenther Benstetter

10. Institute for Microelectronics, TU Wien, Vienna, Austria

    Tibor Grasser

Contributions

K.Z., S.P. and M.L. designed the experiments. X.L. and H.W. provided the CMOS wafers. K.Z., Y.S., Y.Y., W.Z. and O.A. fabricated the hybrid 2D–CMOS microchips. K.Z. and S.P. characterized the microchips. K.Z. measured the logic gates. R.L., H.F. and T.W. assisted in electrical characterization. F.A. did the simulation of the neural network. M.A.V., B.F., A.M. and M.F. assisted in the fabrication, characterization and simulation of structures included in intermediate version of this manuscript that were finally removed. M.M.-R., J.B.R., G.B., X.Z., H.A., T.G. and D.I. gave technical recommendations on fabrication, characterization and/or simulations. M.L. wrote the manuscript, which was revised by all the authors.

Corresponding author

Correspondence to Mario Lanza.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Deji Akinwande, Lain-Jong Li and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 As-received silicon wafers.

Photo of the 200 mm wafer received from the foundry containing all the 2 cm × 2 cm microchips, made of CMOS circuits.

Extended Data Fig. 2 Morphological analysis of the h-BN.

Cross-sectional TEM images of the as-grown multilayer h-BN sheet on Cu. The layered structure is evident, although it contains local defects (i.e., lattice distortions) produced during the CVD process – they are not related to the FIB process because our exfoliated samples never show these features. Those native defects are necessary for the observation of memristive effect. The scale bars (from top to down) and 6 nm, 6 nm, 4 nm and 3 nm.

Extended Data Fig. 3 Nano-chemical characterization of the h-BN stack.

Electron energy loss spectroscopy signal showing the correct structure of the h-BN stack on the CMOS microchip. The image also shows that the titanium electrode can absorb a remarkable amount of oxygen. This oxygen peak is not observed when using gold or silver electrodes.

Extended Data Fig. 4 Electrical characterization of multiple h-BN memristors.

Current versus voltage plots collected for different 1T1M cells using Au–Ti–h-BN–W memristors with a lateral size of 0.053 μm^2. The plots have been drawn with 75% transparency to emphasize the general trend. Each plot contains 50 cycles or more. The first ramped voltage stress is shown in blue colour; it is normally more insulating than the rest, indicating the need of an activation step that softly degrades the h-BN and produces RS. However, no filament is completely formed (see Supplementary Note 1).

Extended Data Fig. 5 Multilevel RS in 1T1M cells.

Current versus time plots for a 1T1M cell when using different end voltage of the reset RVS. Multiple stable conductance levels can be programmed. The composition of the top electrode is Au–Ti. The read voltage applied is 0.1 V. We measured the retention of several states up to 7 h, and the rest up to 200 s due to laboratory availability. The devices show no sign of conductance degradation for any measured state after 7 h, although typical drift (less than 5%) is detected, which is acceptable. According to the Semiconductor Research Corporation [V. Zhirnov, Decadal Plan for Semiconductors: New Trajectories for Memory and Storage, presented in the 2022 Non-Volatile Memory Technology Symposium 2022, December 7th-11, Stanford, USA], the unsaved bytes turnover and lifetime for multiple memory applications is currently less than 15 h, meaning that no retention time above 10 years will be needed. Similarly, the requirements in terms of retention time for memristive electronic synapses is much more relaxed than in most electronic memories, and correct functioning with short retention times of few hours has been claimed [see Kuzum, D., Yu, S., & Wong, H. P. Synaptic electronics: materials, devices and applications. Nanotechnology 24, 382001 (2013)].

Extended Data Fig. 6 Adjustment of R_{LRS}/R_{HRS} in 1T1M cells with Au–Ti top electrodes.

Values of R_{HRS} and R_{LRS} depending on the pulse amplitude of write and erase operation. The value of the R_{LRS}/R_{HRS} window can also be adjusted in this way.

Extended Data Fig. 7 Switching time of 1T1M cells using top Au–Ti electrodes.

Voltage and current versus time plots showing the non-volatile bipolar RS with high temporal resolution. The switching time for the set transition is 232 µs, and the one for the reset transition is 783 ns.

Extended Data Fig. 8 Non-volatile bipolar RS in the 1T1M cells depending on the chemical composition of the top electrode.

Voltage and current versus time plots showing reproducible non-volatile bipolar RS in devices with Au–Ti, Au and Ag top electrode. The operation voltages and switching times are different, providing enough flexibility to operate in different regimes.

Extended Data Fig. 9 In-memory computation with crossbar arrays of h-BN memristors.

a,**b**, Schematic and current versus time plot for “or” and “implication” logic operations (respectively) made with the crossbar array of 1T1M cells in the hybrid 2D/CMOS microchip. The value of the output current changes depending on the conductance of the memristors M1 and M2, which is indicated as “00”, “01”, “10” and “11” on each segment of the plot (“0” and “1” means low and high conductance, respectively).

Extended Data Fig. 10 Non-exhaustive list of materials for microchip manufacturing.

Schematic representation of a microchip; the position of our memristor is indicated with a green star. Our memristors are integrated in the last metallization layer of the BEOL, where Au liners are often employed, and very far from the transistors in the FEOL. Apart from this, Au wire bonding

to the socket is also often employed. Hence, the use of Au electrodes of the memristors is not a concern for this type of devices/circuits. STI means shallow trench isolation.

Supplementary information

Supplementary Information

This file contains Supplementary Figs. 1–14, Tables 1–5 and References.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Zhu, K., Pazos, S., Aguirre, F. *et al.* Hybrid 2D–CMOS microchips for memristive applications. *Nature* **618**, 57–62 (2023). https://doi.org/10.1038/s41586-023-05973-1

- Received: 04 March 2022

- Accepted: 17 March 2023
- Published: 27 March 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05973-1

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

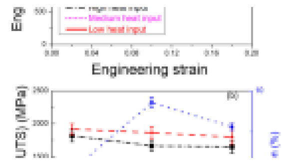
Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **The gap between academia and industry in resistive switching research**
    - Mario Lanza
    - Gabriel Molas
    - Ishai Naveh

  Nature Electronics (2023)

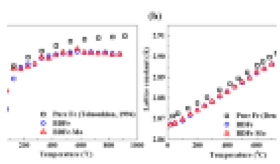
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05973-1



A novel nano-particle strengthened titanium alloy with exceptional specific strength

13 August 2019

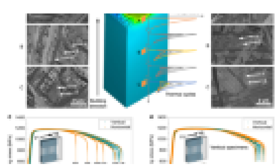
Aniket K. Dutt, Bharat Gwalani, … Raymond E. Brennan



Tailoring grain sizes of the biodegradable iron-based alloys by pre-additive manufacturing microalloying

05 May 2021

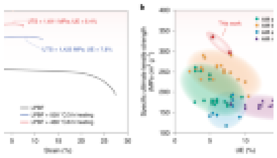
Chih-Chieh Huang, Tu-Ngoc Lam, … E-Wen Huang



Designing against phase and property heterogeneities in additively manufactured titanium alloys

09 August 2022

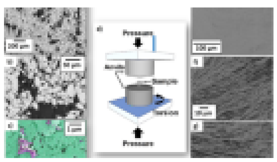
Jingqi Zhang, Yingang Liu, … Ming-Xing Zhang



[**Ultrastrong nanotwinned titanium alloys through additive manufacturing**]()

15 September 2022

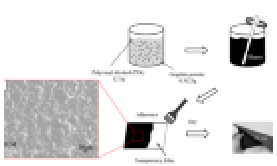
Yuman Zhu, Kun Zhang, … Aijun Huang



[**Titanium-protein nanocomposites as new biomaterials produced by high-pressure torsion**]()

10 January 2023

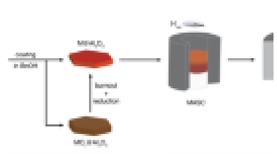
Ricardo Floriano, Kaveh Edalati, … Augusto Ducati Luchessi



[**Fabrication of in-situ rod-like TiC particles dispersed Ti matrix composite using graphite power sheet**]()

09 November 2022

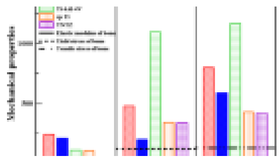
Ning Wang, Yongbum Choi, … Kazuhiro Matsugi



[Tough metal-ceramic composites with multifunctional nacre-like architecture]

15 January 2021

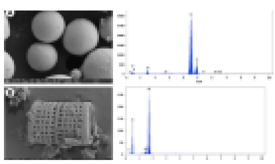
Erik Poloni, Florian Bouville, … André R. Studart



[Functional Studies of Anodic Oxidized β-Ti-28Nb-11Ta-8Zr Alloy for Mechanical, In-vitro and Antibacterial Capability]

24 September 2018

Hsin-I Lin, Yu-Ming Kuo, … Ta-Jen Yen



[Titanium matrix composites reinforced with biogenic filler]

24 May 2022

Izabela Zglobicka, Rafal Zybala, … Krzysztof J. Kurzydlowski

- Article
- Open Access
- Published: 31 May 2023

Strong and ductile titanium–oxygen–iron alloys by additive

manufacturing

- Tingting Song,
- Zibin Chen,
- Xiangyuan Cui,
- Shenglu Lu,
- Hansheng Chen,
- Hao Wang,
- Tony Dong,
- Bailiang Qin,
- Kang Cheung Chan,
- Milan Brandt,
- Xiaozhou Liao,
- Simon P. Ringer &
- …
- Ma Qian

Show authors

Nature **volume 618**, pages 63–68 (2023)

- 3134 Accesses

- 159 Altmetric

- Metrics details

Abstract

Titanium alloys are advanced lightweight materials, indispensable for many critical applications[1,2]. The mainstay of the titanium industry is the α–β titanium alloys, which are formulated through alloying additions that stabilize the α and β phases[3,4,5]. Our work focuses on harnessing two of the most powerful stabilizing elements and strengtheners for α–β titanium alloys, oxygen and iron[1,2,3,4,5], which are readily abundant. However, the embrittling effect of oxygen[6,7], described colloquially as 'the kryptonite to

titanium'[8], and the microsegregation of iron[9] have hindered their combination for the development of strong and ductile α–β titanium–oxygen–iron alloys. Here we integrate alloy design with additive manufacturing (AM) process design to demonstrate a series of titanium–oxygen–iron compositions that exhibit outstanding tensile properties. We explain the atomic-scale origins of these properties using various characterization techniques. The abundance of oxygen and iron and the process simplicity for net-shape or near-net-shape manufacturing by AM make these α–β titanium–oxygen–iron alloys attractive for a diverse range of applications. Furthermore, they offer promise for industrial-scale use of off-grade sponge titanium or sponge titanium–oxygen–iron[10,11], an industrial waste product at present. The economic and environmental potential to reduce the carbon footprint of the energy-intensive sponge titanium production[12] is substantial.

Main

Most industrial titanium (Ti) alloys possess microstructures based on the two basic phases of Ti, the hexagonal close-packed (HCP) α and the body-centred cubic (BCC) β. Represented by Ti–6Al–4V (wt% used throughout unless specified), α–β Ti alloys are the backbone of the Ti industry[1,2]. They can form microstructures comprising[2,3,4,5] (1) lamellar α–β with a near-Burgers orientation relationship, (2) equiaxed α and β or (3) globular α among the α–β lamellae. Each of these microstructures has merits and drawbacks, making α–β Ti alloys versatile for diverse industrial applications[1,2,3,4,5]. Of these, the lamellar α–β microstructure has been commonly applied.

The α–β Ti alloys are formulated by alloying Ti with α-phase and β-phase stabilizers. The α-phase stabilizers are limited to Al, N, O, C, Ga and Ge (refs. [3,4,5]), of which N and C are tightly controlled impurities (0.05% N, 0.08% C)[2,3], whereas Ga and Ge are not commercially viable. Hence, as well as Al, O is the only other practical option. Supplementary Table 1 lists the main α–β Ti alloys using Al as the α-phase stabilizer. Notably, O outshines Al in (1) strengthening the α-phase by a factor of about 20 (calculated according to the data given in Table 4 on page 16 of ref. [1]), (2) stabilizing

the α-phase by a factor of about 10 (based on the aluminium equivalence formula given on page 380 of ref. [5]) and (3) restricting the growth of prior-β grains during solidification by a factor of more than 40 (10.8 versus 0.26)[13]. However, these attributes of O have remained underused in the development of α–β Ti alloys.

The issue with O as a principal α-phase stabilizer in Ti is its embrittling effect owing to its strong interactions with dislocations during deformation[6,7]. Furthermore, O changes the phase equilibria, promoting the formation of the embrittling α_2-phase (Ti_3Al)[14]. These constraints have led to the following empirical design rule for industrial Ti alloys: Al + 10(O + C + 2N) + 1/3Sn + 1/6Zr < 9.0% (ref. [5]). For Ti–6Al–4V, this design rule requires less than 0.12% O (ref. [15]) at 0.05% N and 0.08% C, which was relaxed to 0.13% O for Grade 23 Ti–6Al–4V and 0.20% O for Grade 5 Ti–6Al–4V. Following this rule, a lower Al content allows for a higher O content. Indeed, the latest industrial α–β Ti alloy ATI 425 (Ti–4.5Al–3V–1.8Fe–0.3O)[16], allows 0.3% O maximum because of its lower Al content, for which the above empirical rule accepts a maximum of 0.31% O. If no Al is included, this rule allows a maximum of 0.72% O.

More options exist for β-phase stabilizers in Ti (refs. [3,4,5]), with Fe being the most effective and inexpensive one[17,18]. Furthermore, Fe is the second lightest β-phase stabilizer. However, its use has been constrained by the formation of Fe-stabilized β-flecks during ingot solidification[9] (up to centimetres in size; Supplementary Note 1), which can markedly affect the mechanical properties[9]. Therefore, the use of Fe is usually limited to about 2% in industrial Ti alloys such as ATI 425 and Ti–10V–2Fe–3Al (ref. [2]).

Nonetheless, the aforementioned advantages of O and Fe (≤2%) once attracted substantial efforts to develop α–β Ti–O–Fe alloys as alternatives to Ti–6Al–4V (refs. [19,20,21,22]). The desire for improved hot workability and surface finish during ingot-breakdown hot-working operations was the particular motive at that time[19]. These efforts yielded mixed success, with the two well-investigated compositions being Ti–0.35O–1Fe–0.01N and Ti–0.3O–1Fe–0.04N (refs. [19,20,22]). Both alloys exhibited tensile properties comparable with Ti–6Al–4V in hot-worked and annealed conditions, but

lower tensile strength (600–700 MPa) and ductility (2–3%) in as-cast conditions[20].

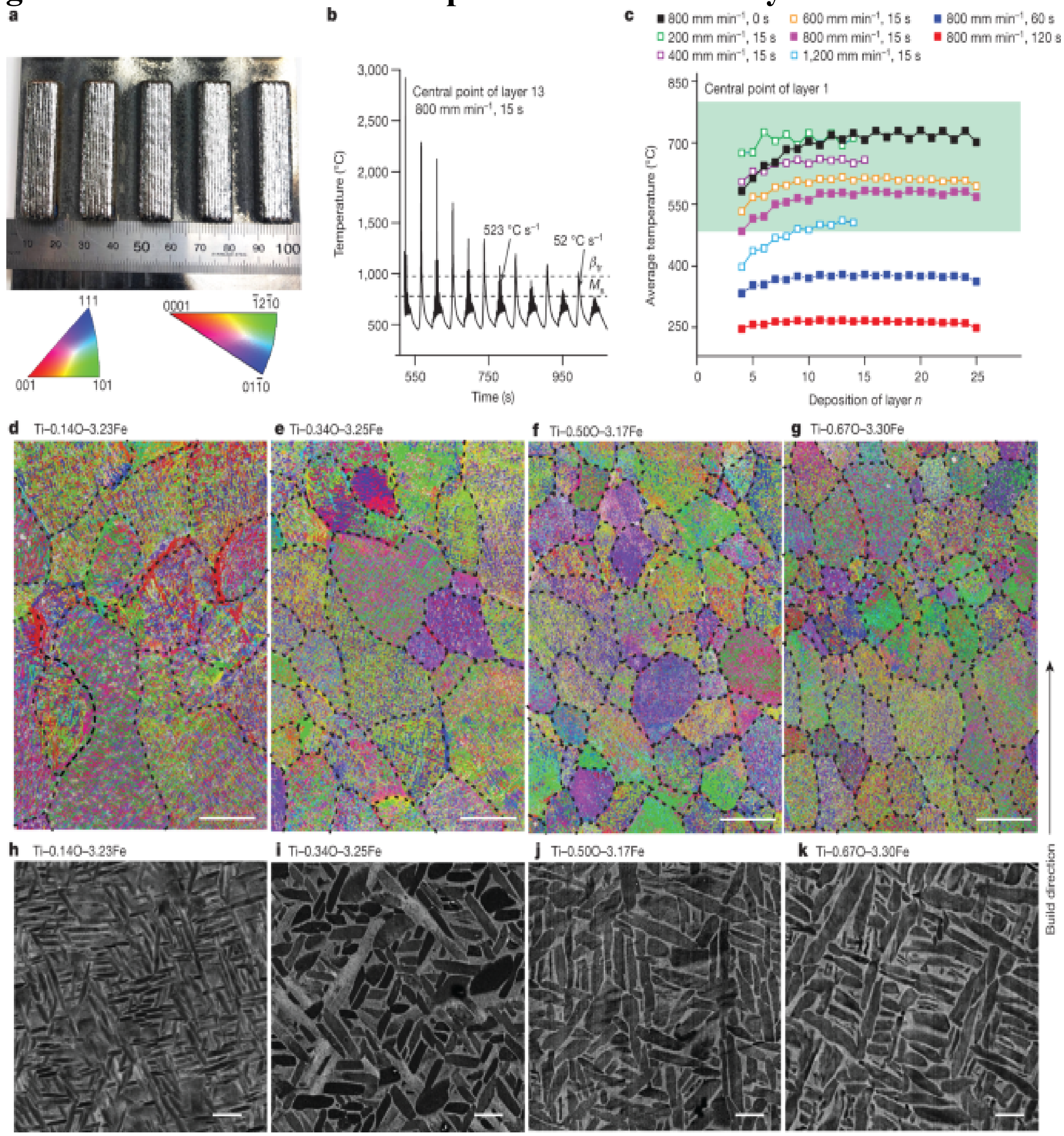
We introduce another consideration of the combined use of O and Fe in Ti, which is related to revitalizing off-grade sponge Ti owing to excess O and Fe contamination[10,11,23]. Sponge Ti production is highly energy-intensive[12]. Given that off-grade sponge Ti (Ti–O–Fe) accounts for 5–10% of all sponge Ti production[10,11,23], their use as input feedstock for powder production for AM has the potential to add notable value and reduce the carbon footprint of the Ti industry.

We have sought to circumvent the metallurgical challenges arising from alloying Ti with O and Fe by integrating alloy design concepts with AM process design. Our goal was to create a new class of ductile and strong α–β Ti–O–Fe alloys through AM. By its molybdenum equivalence, an addition of 3.5% Fe to Ti can retain the prior-β phase to room temperature by water quenching[17,18]. This sets an upper limit for Fe. The maximum O content was set at 0.7% based on the empirical rule discussed earlier (0.72% O). Furthermore, most Ti alloys lose tensile ductility above 0.7% O (based on Fig. 1 on page 6 of ref. [1]). Four levels of O (0.15%, 0.35%, 0.50%, 0.70%) were thus considered with 3% Fe, leading to ten experimental alloys (Extended Data Table 1). In terms of AM process selection, we chose laser metal powder directed energy deposition (DED), which—aided by high-fidelity simulations—allows for the fabrication of large-scale near-net-shape components with a consistent microstructure.

We first simulated DED of rectangular coupons (Fig. 1a), using parameters from Extended Data Table 2. The simulations predicted high cooling rates after solidification (Fig. 1b), which are expected to lead to metastable phases such as α′-martensite. However, the large number of thermal pulses (Fig. 1b), high stabilization temperatures (Fig. 1c) and approximately isothermal durations should ensure that only α and β phases are present. For this purpose, we chose 480–800 °C as the required thermal history bounds (Methods), delimited as the green zone in Fig. 1c. Rectangular coupons of the designed compositions (Extended Data Table 1) were then printed within and outside this window, together with two reference alloys, Ti–6Al–4V–

0.22O–0.20Fe and Ti–0.16O–0.062Fe. The printed coupons exhibited consistent compositions (Supplementary Note 2).

Fig. 1: Microstructure of DED-printed Ti–O–Fe alloys.



a, As-built rectangular coupons of 40 × 10 × 5 mm^3 at the layer thickness of 200 µm. The five coupons in **a** all have the same composition of Ti–0.34O–3.25Fe. **b**, Temperature profile of the central point of layer 13 in a 25-layer coupon by simulation. **c**, Processing window (green zone) determined by simulation. **d**–**k**, Electron backscatter diffraction (EBSD) inverse pole figure

images (**d**–**g**; scale bars, 100 μm) and backscattered electron images (**h**–**k**; scale bars, 1 μm) of the printed Ti–0.14O–3.23Fe (**d**,**h**), Ti–0.34O–3.25Fe (**e**,**i**), Ti–0.50O–3.17Fe (**f**,**j**) and Ti–0.67O–3.30Fe (**g**,**k**) alloys. The light-grey phase in **h**–**k** is the β phase and this contrast arises owing to Fe enrichment.

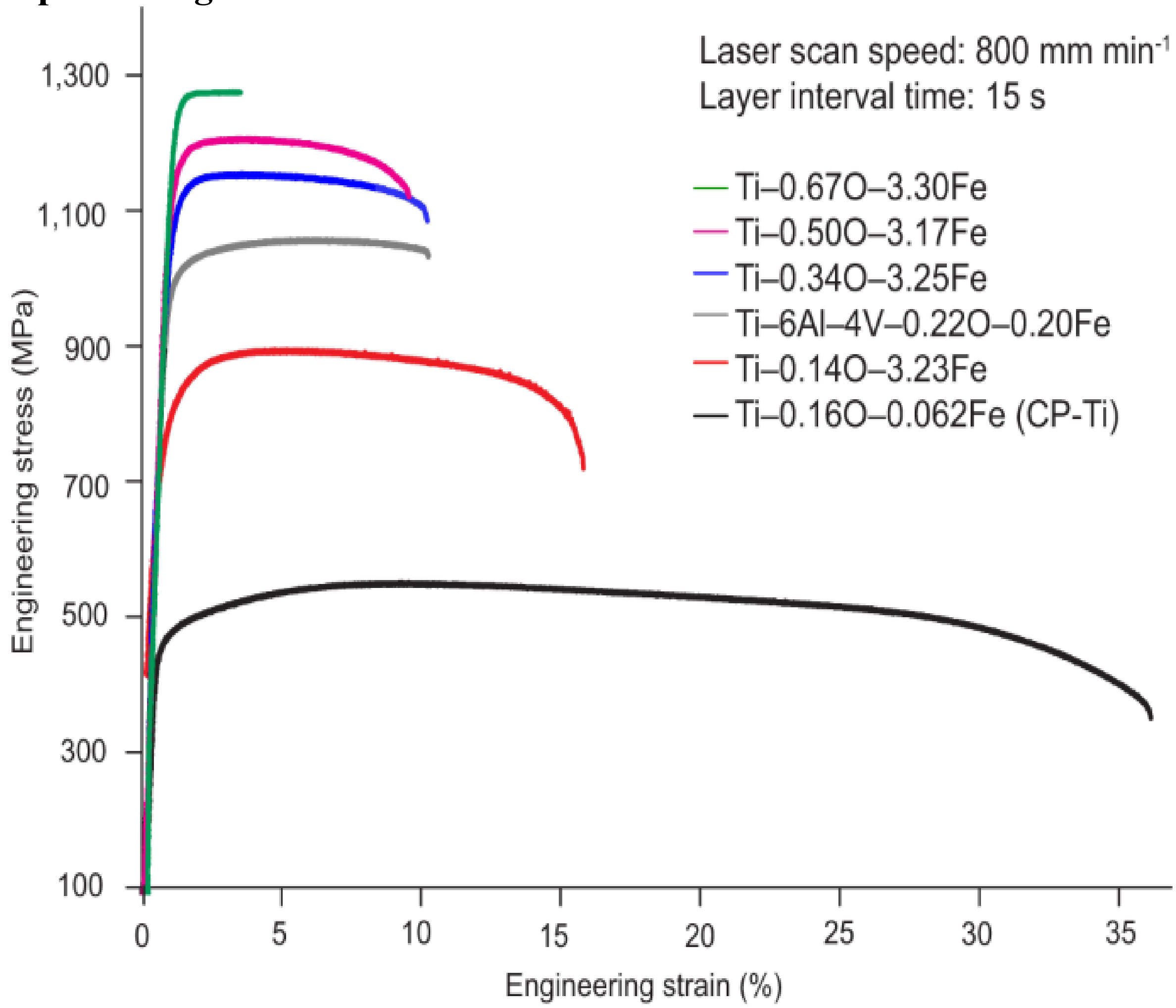
Source data

The microstructure of the low-oxygen Ti–0.14O–3.23Fe alloy comprises short columnar and equiaxed prior-β grains (Fig. 1d), akin to the laser additively manufactured Ti–6Al–4V–3Fe alloy[24]. Fine equiaxed prior-β grains formed with increasing O content (Fig. 1e–g and Supplementary Fig. 1) compared with long columnar prior-β grains in Ti–6Al–4V printed under similar[25] or different conditions[26,27,28]. At room temperature, fine α–β lamellae were the prevailing microstructure in each alloy (Fig. 1h–k). The α-lath thickness increased from 180 ± 33 nm at 0.14% O to 375 ± 76 nm at 0.67% O, whereas the β-phase volume fraction simultaneously increased from 21 ± 2.3% to 31 ± 0.5% (Supplementary Fig. 2). No Fe-stabilized β-flecks were observed in any of these printed alloys, but they were prevalent in the copper-mould-cast Ti–0.35O–3Fe alloy (Extended Data Fig. 1). The reasons are discussed in Supplementary Note 1. The avoidance of β-flecks is an important advantage of AM in the fabrication of these alloys (Supplementary Note 3).

Figure 2 shows representative tensile properties of four designed Ti–O–Fe alloys and two reference alloys. Our goal here was to change alloy composition without changing the AM processing conditions, within the required thermal history bounds (Fig. 1c). The complete engineering stress–strain curves are provided in Extended Data Fig. 2a. Separate coupons were printed by fixing the alloy composition to Ti–0.35O–3Fe and varying the AM processing condition. The tensile stress–strain curves of these alloys and the copper-mould-cast Ti–0.35O–3Fe alloy are shown in Extended Data Fig. 2b,c, whereas Extended Data Fig. 3 shows their microstructures. Without optimization, our Ti–(0.34–0.50)O–(3.17–3.32)Fe alloys printed within the processing window exhibited tensile ductility (ε_f) from 9.0 ± 0.5% to 21.9 ± 2.2% (the change in ε_f is not because of porosity; Supplementary Note 2) and ultimate tensile strength (σ_{UTS}) from 1,034 ± 9 to 1,194 ± 8 MPa

(Extended Data Table 1). The as-cast Ti–0.35O–3Fe alloy demonstrated a more than 50% lower ε_f at similar σ_{UTS}.

Fig. 2: Tensile properties of DED-printed Ti–O–Fe alloys at room temperature by focusing on varying alloy composition without changing the processing conditions.



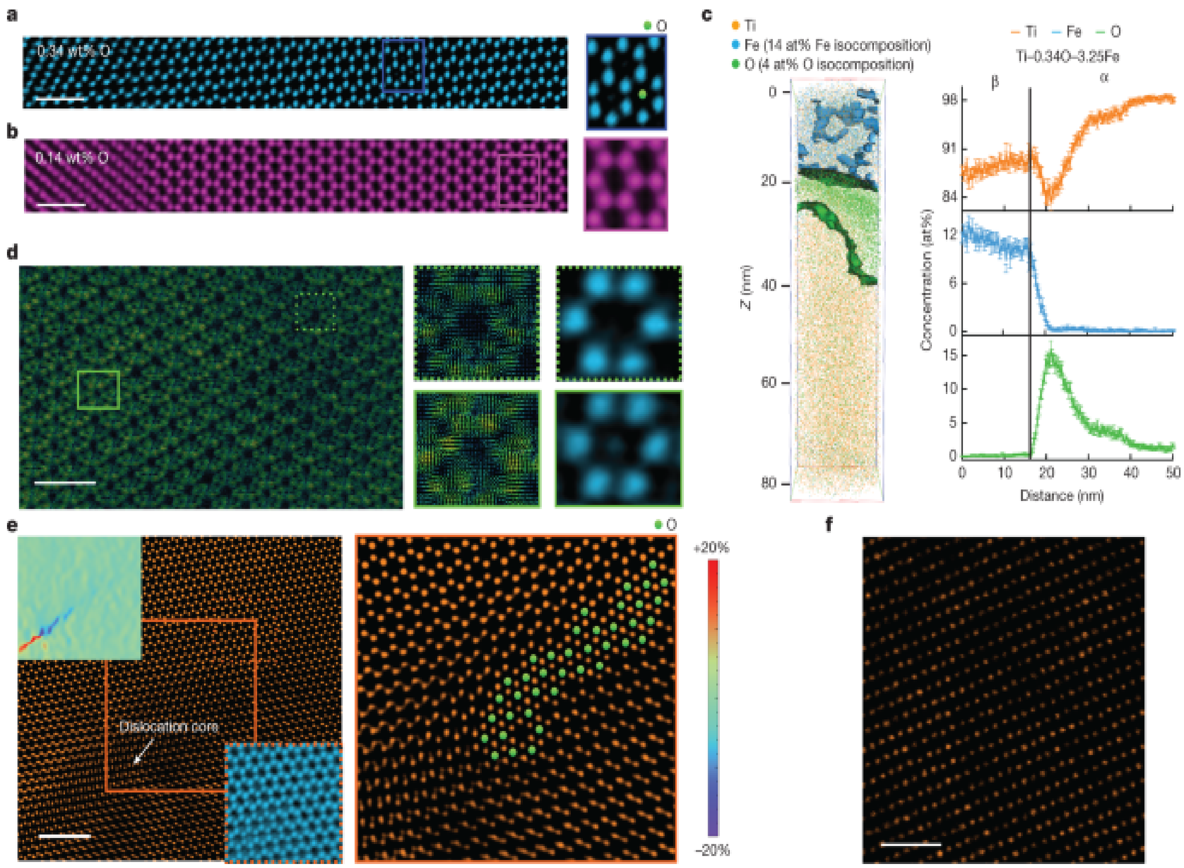
Engineering stress–strain curves of the Ti–(0.14–0.67)O–(3.17–3.30)Fe, ultralow-iron Ti–0.16O–0.062Fe and Ti–6Al–4V alloys, printed using the same DED conditions (laser scan speed: 800 mm min^{-1}, layer interval: 15 s; see Extended Data Table 2 for other conditions). Extended Data Table 1 lists the tensile properties of each alloy.

The strengthening potency of the Fe reached 105 MPa/1.0 wt% Fe by comparing the σ_{UTS} values of the Ti–0.14O–3.23Fe and Ti–0.16O–0.062Fe

alloys, close to the reported experimental value of 75 MPa/1.0 wt% Fe (based on the data given in Table 12 on page 29 of ref. [1]). The strengthening potency of the O registered 76 MPa/0.1 wt% O by comparing the σ_{UTS} values of the Ti–0.14O–3.23Fe and Ti–0.67O–3.30Fe alloys, which is within the reported typical experimental range of >70 MPa/0.1 wt% O (refs. [29,30]). Both Fe and O played an important role in strengthening these alloys.

To explain the strengthening mechanisms, we investigated the atomic distribution of the O and Fe in three alloys, Ti–0.14O–3.23Fe, Ti–0.34O–3.25Fe and Ti–6Al–4V–0.22O–0.20Fe. Combined with integrated differential phase contrast (iDPC), scanning transmission electron microscopy (STEM) enables direct observations of interstitial light elements[31]. We observed strong segregation of O atoms to the α-lath rims in the Ti–0.34O–3.25Fe alloy (Fig. 3a). They existed in the HCP interstitial sites near the α/β interface, forming a unique nano-heterogeneity from the interior of the α-lath (low O, ductile) to the interface regions (high O, strong). By contrast, this nano-heteromicrostructure was rarely observed in the low-oxygen Ti–0.14O–3.23Fe alloy (Fig. 3b). Atom probe tomography (APT) analysis confirmed the same observations (Fig. 3c and Extended Data Fig. 4a).

Fig. 3: Distribution of O and Fe atoms in DED-printed α–β Ti–O–Fe alloys.



a,**b**, iDPC-STEM images of the α/β interfaces in Ti–0.34O–3.25Fe and Ti–0.14O–3.23Fe. Oxygen atoms are observed at the interstitial positions in the α-phase near the α/β interface in Ti–0.34O–3.25Fe, whereas few such O atoms are detected in the Ti–0.14O–3.23Fe alloy. Scale bars, 1 nm. **c**, APT data from Ti–0.34O–3.25Fe, highlighting the tendency of O atoms to segregate towards the edges of the α-phase near the α/β interfaces. **d**, Tensor-flow DPC-STEM images along a $[0001]_\alpha$ direction. The tensor direction represents the local electric field direction. The tensor colours represent the strength of the local electric field (green = weaker, yellow = stronger). Scale bar, 1 nm. **e**, HAADF-STEM image of a dislocation inhibited by an O interstitial array. Geographic phase analysis (top-left corner inset) shows the strain condition of the defect and the surrounding region. The dislocation core is defined by the red–blue intersection in the inset. The iDPC-STEM image (bottom-right corner inset) shows the O interstitial array. The right-hand enlarged HAADF-STEM image shows the strong presence of O interstitials (extracted from the iDPC image and marked as green balls) around the dislocation, impeding dislocation movement. Scale bar, 2 nm. **f**, An HAADF-STEM image of a β-phase

region, highlighting the non-uniform distribution of Fe, shown by the uneven *Z*-contrast, for which the zone axis is $[110]_\beta$. Bright contrast means more Fe atoms in the local β-Ti lattice. Scale bar, 1 nm.

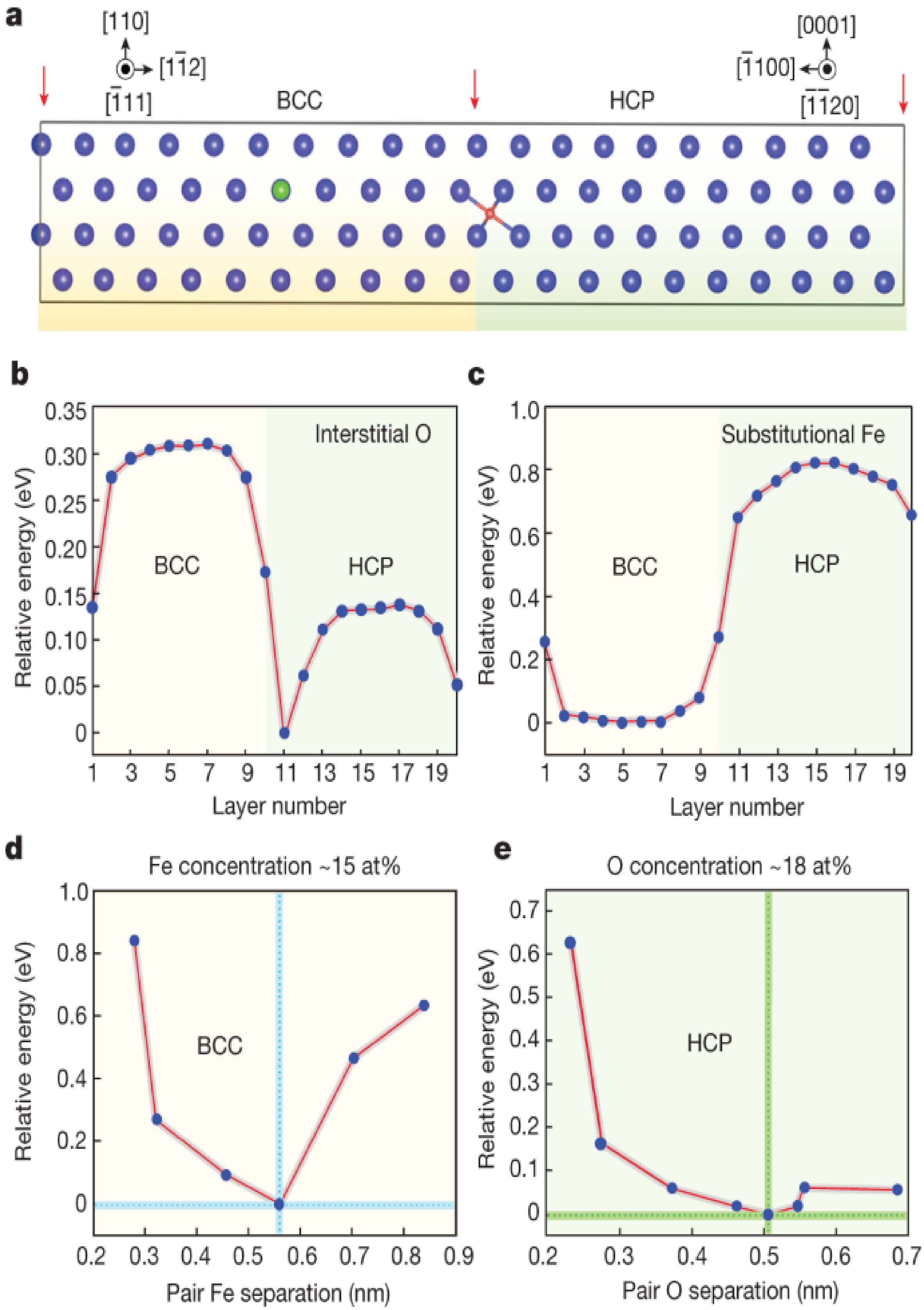
The local charge density in the Ti–0.34O–3.25Fe alloy was analysed using differential phase contrast (DPC)STEM (Fig. 3d). Here the vectors represent the local electric-field directions, whereas the colours denote the relative strength of the local electric field (green = weak, yellow = strong). An increase in the local charge density (yellow) was observed in the α-phase rims in which O interstitials had segregated (Fig. 3d). This is indicative of a further bonding contribution in the lattice from the O interstitials, which we suggest contributes to strengthening by further impeding dislocation motion. Figure 3e shows a collection of O-interstitials at a dislocation in an α-phase rim, similar to Cottrell atmospheres in BCC crystals[32], recorded using combined iDPC and high-angle annular dark-field (HAADF)-STEM techniques. The dislocation is effectively pinned by such O atmospheres.

Our APT data further showed that the β-phase was virtually free of O (0.03 at%). This was supported by the CALPHAD (CALculation of PHAse Diagrams) predictions performed at 650 °C in the green zone of Fig. 1c (Supplementary Table 2). By comparison, the β-phase in the printed Ti–6Al–4V–0.22O–0.20Fe alloy contained substantial O (0.27 at%) (Extended Data Fig. 4b,c), consistent with the literature (Supplementary Table 4). Our APT data also showed that the α-phase was virtually free of Fe (0.02 at%; Fig. 3c), whereas the substantial Fe in the β-phase was non-uniformly distributed (Fig. 3f), which adds a strengthening capacity to the alloy by generating local strains for impeding dislocation motion.

Density functional theory (DFT) simulations (Fig. 4a) predict that the O atoms prefer to reside in the α-phase, especially close to the α/β interface (Fig. 4b), but the Fe atoms show no tendency to segregate to the interface (Fig. 4c). These predictions fully support the observations in Fig. 3. Furthermore, the 4th and 3rd nearest-neighbour separation configurations were most favourable for Fe–Fe pairs in the β-phase (Fig. 4d), whereas the 5th nearest neighbour ($d = 0.506$ nm) separation configuration was most favourable for O–O pairs in the α-phase (Fig. 4e). On this basis, we estimated the [Fe] and [O] content in each phase, which gives 15 at% Fe in β

and 18 at% O in α. These estimates match our APT measurements well (Fig. 3c).

Fig. 4: DFT simulations of the distribution of Fe and O atoms in BCC (β) and HCP (α) phases of α–β Ti–O–Fe alloys.



a
[110]
[1$\bar{1}$2]
[$\bar{1}$11]
BCC
HCP
[0001]
[$\bar{1}$100]
[$\bar{1}\bar{1}$20]
b
Interstitial O
BCC
HCP
Relative energy (eV)
Layer number
c
Substitutional Fe
BCC
HCP
Relative energy (eV)
Layer number
d
Fe concentration ~15 at%
BCC
Relative energy (eV)
Pair Fe separation (nm)
e
O concentration ~18 at%
HCP
Relative energy (eV)
Pair O separation (nm)

a, A DFT BCC ($\bar{1}1\bar{2}$)/HCP ($\bar{1}100$) interface model containing one octahedral interstitial O atom (in red) and one substitutional Fe atom (in green) at their respective favourable positions. On the basis of Fig. 3, we constructed a 1 × 1 in-plane ten-layer BCC ($\bar{1}1\bar{2}$)/1 × 2 in-plane ten-layer HCP ($\bar{1}100$) interface structure (ten layers on each side) to simulate an α/β interface. The red arrows indicate the equivalent BCC/HCP interfaces owing to the periodic boundary conditions. **b**,**c**, The calculated layer-resolved relative energy mapping for an interstitial O atom (**b**) and a substitutional Fe atom (**c**) throughout the interfaces. The relative energy as a function of pair separation for O and Fe in each phase was also calculated. **d**,**e**, The calculated relative total energy as a function of pair separation for pair interstitial O atoms in a 96-atom HCP-Ti supercell (**d**) and pair substitutional Fe atoms in a 54-atom BCC-Ti supercell (**e**). The calculated equilibrium O and Fe concentrations were based on the respective energetically most favourable configurations (see text).

Source data

We also sought a direct comparison of the crystallographic arrangements of the α-variants in our printed Ti–(0.14–0.67)O–(3.17–3.30)Fe and Ti–6Al–4V alloys. The prevalent misorientation between neighbouring α-variants in each Ti–O–Fe alloy was $\left[11\bar{2}0\right]/60^\circ$, compared with $[\bar{10}55\bar{3}]/{63.26}^{\circ}$ in Ti–6Al–4V (Extended Data Fig. 5). The former is close to $\{10\bar{1}1\}$ twinning[33], possessing lower energy than the latter, and is therefore an energetically preferred configuration[34,35].

Dislocation activity in the vicinity of tensile crack tips determines whether materials exhibit brittle or ductile behaviour when fracture occurs[36]. Active dislocation multiplication can relieve stress concentration and blunt crack tips, enabling further plastic deformation[37]. Our α–β Ti–(0.35–0.70)O–3Fe alloys comprise roughly 30 vol% β-phase versus about 5 vol% β-phase in Ti–6Al–4V. We suggest that this large volume fraction of the virtually oxygen-free β-phase (Fig. 3c) and the oxygen nano-heterogeneity inside the α-phase have combined to play a vital role in mediating the overall deformation process, resulting in excellent tensile properties. Therefore, we

examined areas abutting the tensile fracture surfaces of the Ti–0.34O–3.25Fe ($\varepsilon_f = 9.0\%$) and Ti–0.67O–3.30Fe ($\varepsilon_f = 3.0\%$) alloys using TEM.

Dislocation tangles developed at the centre of the α-phase laths in the Ti–0.34O–3.25Fe alloy, whereas extensive dislocation multiplication was apparent in the β-phase (Extended Data Fig. 6). This provides direct evidence to support our above hypothesis. In the Ti–0.67O–3.30Fe alloy, dislocation multiplication was observed, but to a lesser extent (Extended Data Fig. 7), corresponding to the limited overall plastic deformation ($\varepsilon_f =$ 3.0%). The tendency towards fracture increases with increasing applied tensile stress (σ) until the local stress intensity factor K (proportional to σ) exceeds the critical value for the α-phase or the prior-β grain boundaries or the interfacial regions between the high-oxygen α-lath rim and β-phase. A local fracture would then proceed through one or all of them, for example, when σ reached 1,271 ± 6 MPa for the Ti–0.67O–3.30Fe alloy, leading to low ε_f.

The above understanding corresponds well to the fracture features of each alloy (Extended Data Fig. 8a–f). For example, the fracture surface of the Ti–0.34O–3.25Fe alloy consisted of large, deep dimples (extensive dislocation multiplication) and a portion of facets with small, shallow dimples (less ductile). Conversely, the fracture surface of the Ti–0.67O–3.30Fe alloy exhibited notable facets with a size comparable with the prior-β grain size, indicative of inter-(prior-β)-granular fracture, and only a fraction of large, deep dimples (localized dislocation multiplication). The fracture characteristics of the other alloys could be understood similarly.

The relative phase fractions and grain sizes also play an important role in the strengthening mechanisms. The increase in the β-phase volume fraction with increasing O stems from the strong partitioning of O and Fe in each phase. For instance, at 800 °C (in the green zone of Fig. 1c), CALPHAD predicts substantially more β-phase in the Ti–0.14O–3Fe alloy (53.5 vol%) than in the Ti–0.67O–3Fe alloy (39.5 vol%) (Supplementary Note 5). The lower the β-phase fraction, the higher the average Fe content in the β-phase and the more stable the β-phase will be. This leads to more retained β-phase. In this particular case, the β-phase in the Ti–0.67O–3Fe alloy contained 7.30% Fe versus 4.85% Fe in the β-phase of the Ti–0.14O–3Fe alloy. In terms of the

increase in the α-lath thickness with increasing O, we ascribe this to the higher formation temperature of the α-phase, which accelerates diffusion while reducing strain energy from the formation of the α-phase.

To summarize, we have demonstrated an integration between alloy design and simulation-based AM process design to create a new class of strong and ductile α–β Ti–(0.35–0.50)O–3Fe alloys (ε_f = 9.0 ± 0.5% to 21.9 ± 2.2%; σ_{UTS} = 1,034 ± 9 to 1,194 ± 8 MPa), available across a generous AM processing window, using the readily abundant elements O and Fe. We attribute the success of these alloys to a combination of multiscale microstructural features arising from this integration. These include: (1) the fine α–β lamellae distributed within the fine equiaxed prior-β grains; (2) the high potency of O and Fe in strengthening the α-phase (virtually Fe-free) and β-phase (about 30 vol%, virtually O-free), respectively; and (3) the preferred misorientation configuration between the neighbouring α-variants in these alloys. Among these factors, the unique partitioning of O and Fe is fundamental, and of particular criticality is our report of a nano-heteropartitioning in the α-phase leading to high-oxygen (strong) and low-oxygen (ductile) distributions that affect the local nature of the atomic bonding, shown by the DPC and iDPC.

These strong and ductile Ti–O–Fe alloys are expected to have implications across a wide range of potential applications at room temperature (Supplementary Note 6). Furthermore, sponge zirconium (Zr) is produced in the same way as sponge Ti. Therefore, the same could be expected for the use of off-grade sponge Zr to develop strong and ductile Zr–O–Fe alloys. In addition, this work provides a potential pathway for future interstitial engineering by AM, such as mitigating nitrogen (N) embrittlement in Ti and Zr, and oxygen embrittlement in other metals (Supplementary Note 6).

Methods

Materials for alloy fabrication

The feedstock powders used for alloy fabrication through laser DED include argon-gas-atomized commercially pure Ti (CP-Ti) powder (50–100 µm, TLS Technik GmbH & Co.), water-atomized Fe powder (20–50 µm, Höganäs)

and TiO_2 powder (<5 μm, Sigma-Aldrich). The composition of the CP-Ti powder is Ti–0.14O–0.139Fe–0.01N–0.011C–0.0011H and that of the Fe powder is Fe–0.003C–0.09O–0.01S (in wt%). The powders were blended in a Turbula mixer (model T2F) as per each designed alloy composition for 2 h at ambient temperature in a sealed plastic container. For comparison purposes, powders of an ultralow-iron CP-Ti composition (Ti–0.16O–0.062Fe) and a Ti–6Al–4V–0.22O–0.20Fe alloy were also used to build reference samples. Both powders were supplied by TLS Technik GmbH & Co. with a particle size range of 50–100 μm.

Furthermore, CP-Ti (Ti–0.13O–0.15Fe) and pure Fe nuggets (99.99%, ZhongNuo Advanced Material Technology Co.) and TiO_2 powder (<5 μm, Aladdin) were used for alloy fabrication by means of vacuum arc melting and casting.

Thermal history simulation

The DED module in Simufact Welding was used to track temperature evolution in the build[25,38]. The predictability of Simufact Welding (DED) in terms of melt pool shape and size and thermal cooling was evaluated in Supplementary Note 2. We used the temperature-dependent thermophysical data for Ti–6Al–4V owing to the lack of similar data for Ti–O–Fe alloys. The simulation focuses on the deposition of a 25-layer rectangular coupon ($40 \times 10 \times 5$ mm^3) on a 10-mm-thick Ti–6Al–4V substrate.

The DED module adopts a finite element method without considering melt pool convections but focusing on heat conduction[38]. This is a common simplification in weld modelling because both the filler metal and the interface scale are negligible compared with the base metal dimensions[39]. In the case of laser metal powder DED, the melt pool size is substantially smaller than the substrate or part size so that most of the heat is conducted away from the melt pool towards the substrate.

To use the DED module, the deposit geometry is first introduced as a defined precursor, the meshes of which are set to the quiet mode (inactive) at the outset. Once the simulation starts, each mesh is activated (element-wise) by means of the moving volumetric heat source entity. The elements that are

touched by the heat source during its propagation will be permanently activated using an element-birth algorithm. As a result, all the activated elements are exposed to the heat source, which represents a 3D volumetric heat flux that has the dimensions of a typical melt pool. Being a nodal load, the heat flux from this heat source will increase the temperature of the just-activated elements to values above their melting or liquidus temperature.

Extended Data Table 2 lists the DED parameters and illustrates the bidirectional scan strategy used for simulation, including an equivalent laser spot size of 1.5 mm and a laser power of 500 W. The selection of DED parameters is discussed in the next section. The scanning path files (G-code) with the DED parameters are imported into the DED module to build the part geometry. Each layer (40 × 10 mm) consists of 6,400 points with a resolution of 0.25 × 0.25 mm. The time resolution is set to 0.02 s to capture heating and cooling information. The simulation focuses on the influence of the layer-to-layer interval (0 s, 15 s, 60 s and 120 s).

The multipeaks or thermal pulses in Fig. 1b and Supplementary Fig. 3 arise from the bidirectional scan strategy used. Each odd-number layer has nine scan paths parallel to the length direction of the sample across its width (10 mm). As a result, nine peaks or nine thermal pulses corresponding to these nine scan paths will appear in previously solidified layers. By contrast, each even-number layer has 38 short scan paths perpendicular to the length direction of the sample. However, each of these short scan paths is not thermally strong enough to result in a detectable thermal pulse in the previously solidified layers. Collectively, they only lead to a limited number of small thermal pulses. Supplementary Fig. 3 shows the correspondence in detail.

Alloy fabrication by DED

Four groups of α–β Ti–O–Fe alloys were designed: Ti–0.15O–3Fe, Ti–0.35O–3Fe, Ti–0.50O–3Fe and Ti–0.70O–3Fe. Extended Data Table 1 lists the measured compositions of each fabricated alloy in the range of Ti–(0.14–0.67)O–(3.11–3.36)Fe. They were all fabricated using a TRUMPF TruLaser Cell 7020 system (thin-disk laser, 1,030 nm). A 12-mm-thick CP-Ti substrate plate was used for deposition. Extended Data Table 2 lists the experimental deposition parameters. As-built coupons (length: 40 mm,

width: 10 mm, thickness: 5 mm) are shown in Fig. 1a and Supplementary Fig. 5a. Five coupons of each composition were printed.

Selection of DED parameters and processing window

The parameters selected should ensure fast and consistent high-quality fusion (virtually free of lack-of-fusion defects and keyhole pores) and result in fine and short α–β lamellae through decomposition of the α′-martensite phase after solidification. The laser energy density E_d (J mm^{-2}) for obtaining high-quality DED of Ti–6Al–4V on a Ti–6Al–4V alloy substrate has been systematically studied in the range 16.6–36 J mm^{-2} with a large laser spot size (2 mm)[40], in which \({E}_{{\rm{d}}}=\frac{P}{dv}\) (*P* laser power, *d* laser spot site, *v* laser scan speed). The use of high values of E_d was not recommended[40]. This is consistent with our experiences of using the TRUMPF system for laser metal powder DED of Ti–6Al–4V on Ti–6Al–4V substrates with a similar laser spot size (optimum range of E_d = 25–35 J mm^{-2}; laser spot size: 1.5 mm).

The spot size was chosen to be 1.5 mm to ensure a sufficiently large melt pool (≥1.0 mm^3), which is important for (1) enhancing chemical homogeneity when mixed powders are used, (2) layer height uniformity and (3) achieving a more stable thermal profile in the build. A wide range of scan speed (200–2,000 mm min^{-1}) is available. However, with E_d = 25–35 J mm^{-2}, our experience with the TRUMPF system is that, for high-quality fusion with the previously solidified layers, the laser spot dwell time *d*/*v* should be 0.1–0.15 s. This leads to the selection of *v* = 600–800 mm min^{-1} and, therefore, the laser power of 500 W for the expected E_d.

The layer thickness of 200 μm and the powder flow rate of 1.7 g min^{-1} were determined together through various previous experimental studies to obtain (1) a consistent weld bead with an aspect ratio of 3–4 to 1, (2) a high-quality build that is essentially free of lack-of-fusion defects (through fracture surface characterization, assisted with microcomputed tomography analysis) and (3) a consistent microstructure.

With the above-selected parameters, the DED module in Simufact Welding predicted a semiellipsoidal melt pool with a volume of 1.08 mm^3 (major axis: 2.3 mm; minor axis: 1.5 mm; depth: 0.6 mm; see Supplementary Note 2), in which the liquidus temperature of our designed Ti–0.35O–3Fe alloy is about 1,659 °C.

The scan spacing was selected as 1.05 mm based on the simulated melt pool width of 1.5 mm to ensure a linear overlap of 70% for sufficient fusion between two abutting tracks, a standard practice to achieve layer and height uniformity during the build (an overlap of 30% minimum has been commonly used).

The layer-to-layer interval time (τ) can be varied over a wide range. A practical range is from 0 s to no longer than 120 s. Changing τ can substantially change the cooling rate and the entire thermal history of the build, thereby altering the microstructure. The default layer interval used in this work was 15 s but we have investigated the whole range of 0–120 s.

These preliminary selected parameters were then subjected to systematic simulations using Simufact Welding to meet the following two criteria to obtain α–β lamellae through decomposition of the α′-martensite: (1) an appropriate cooling rate ($\geq$ 400 °C s^{-1}) when the β-phase in each solidified layer is cooled from the single β-phase region for the formation of α′ and (2) an appropriate stabilized temperature window across each layer ($\geq$ 480 °C, preferably 600–750 °C) (excluding the last few layers) to allow α′ → α–β.

If the selected parameters fail to satisfy these two criteria, reselection by simulation is necessary until they do. Supplementary Table 3 summarizes the predicted average cooling rate at three selected points in the build before reaching 800 °C (the assumed martensite start temperature, M_s), with the scan speed ranging from 200 to 1,200 mm min^{-1} and the layer interval ranging from 0 to 120 s. They all satisfy the required cooling condition.

Figure 1c shows the evolution of the average temperature at the centre of layer 1 from all 24 subsequent layers of deposition. The average temperature at the central point was calculated over the scanning time for each subsequent layer (for example, 35.85 s at 600 mm min^{-1} for each odd-

number layer) plus the layer interval time (for example, 15 s). Stabilization occurs from the 8th or 9th layer of deposition.

Decomposition of α′ starts to occur at 400 °C (long holding is needed)[41] but complete decomposition of α′ requires isothermal exposure to 800 °C for 2 h or longer[41,42]. However, the effect of thermal pulses (Fig. 1b and Supplementary Fig. 3) introduced during DED can markedly accelerate the decomposition process[25].

On the other hand, the microstructure could risk severe coarsening at ≥800 °C, whereas below 480 °C (the minimum stress relief temperature), decomposition of α′ may not be complete. Furthermore, the stabilized temperature increases with increasing build height (Supplementary Fig. 4), which means that the stabilized temperature of the first layer should not be set too high. On the basis of these considerations, we choose 480–800 °C in Fig. 1c (the green zone) as the required thermal history bounds or processing window. The stabilized temperature increases with reducing layer interval time or scan speed or with increasing build height.

The scan speed (600–800 mm min^{-1}) and layer interval time (0–120 s) were then varied both within and outside the simulation-informed processing window to investigate the resulting microstructures and tensile properties of our designed Ti–(0.15–0.70)O–3Fe alloys.

Chemical homogeneity of as-fabricated samples

Each melt pool produced in the deposition process of this work (around 1 mm^3) requires melting of about 6,250 Ti powder particles used in this study ($D_v50 = 69.09$ μm). This helps to mitigate the chemical inhomogeneity issue arising from the use of mixed powders. Also, because the melt pool depth (0.6 mm) is three times the thickness of each solidified layer (0.2 mm), and there is a 70% overlap between each two abutting tracks as well, substantial remelting occurs to each solidified melt pool. This further facilitates chemical homogeneity. Both compositional analysis and microstructural examination of different batches of as-fabricated coupons confirmed good overall chemical homogeneity (see Supplementary Note 2).

Alloy fabrication by arc melting and water-cooled copper-mould casting

As-cast coupons of the Ti–0.33O–3.11Fe alloy were prepared using a vacuum arc melter equipped with a suction casting device (Physcience Opto-electronics Co., Ltd., model WK-II type). The water-cooled copper (Cu) hearth in the melting chamber has five hemispherical cavities (cleaned before use). The TiO_2 powder was placed at the bottom of one cavity, followed by the properly cleaned CP-Ti and Fe nuggets to avoid powder splashing during arc melting. Furthermore, separate CP-Ti nuggets were placed in another cavity as oxygen scavenger. The furnace was first evacuated to 5×10^{-3} Pa and then purged with argon to 0.5 Pa. The vacuum and purge processes were repeated three times to ensure a low oxygen level chamber.

The scavenging CP-Ti nuggets were first melted three times for oxygen scavenging. Then the alloy charge (45–55 g) was melted and remelted eight times in total. The alloy button ingot was flipped over after each melting for chemical homogenization. Then it was moved to the central cavity of the water-cooled Cu hearth using a small shovel in the chamber. The alloy was remelted again and kept in the molten state for 30 s. The suction casting device (a water-cooled Cu mould) was connected to the melting chamber. After remelting, an air pressure difference was introduced between the upper and lower parts of the facility by opening the air evaporation valve connected to the Cu mould cavity. The molten alloy was then air-sucked and solidified in a water-cooled Cu mould cavity (length: 120 mm, width: 12 mm, thickness: 5 mm). Five coupons were fabricated following the same process.

Microstructural characterization

EBSD was conducted on a JEOL JSM-7200F scanning electron microscope (SEM) system with an Oxford Instruments AZtecHKL imaging system, operated at 20 kV. Each specimen was tilted at 70° and the step size used was 500 nm. The EBSD data were analysed using Channel 5 software. Microstructure and fracture surface analyses were performed using an SEM (FEI Verios 460L) in either the secondary electron or backscattered electron

mode. We conducted quantitative image analysis using the Image-Pro Plus package for the measurement of the width of α-laths and areas of the α and β phases for the volume fraction of β. STEM observations were carried out using a Themis Z double-corrected microscope operated at 300 kV equipped with an FEI DF4 detector for DPC and iDPC imaging.

The APT characterization was performed using a laser-assisted CAMECA local electrode atom probe 4000X Si. The needle-shaped specimens were prepared by the lift-out method and annual-milled in a traditional Ga FIB-SEM (Zeiss Auriga) and a plasma FIB-SEM (Helios G4). The specimens were analysed in laser mode with a specimen temperature of 50 K, a laser energy of 50 pJ and a pulse frequency of 200 kHz. Visualization and Analysis Software version 3.8.4 was used for the 3D reconstruction and data analyses. The default values of detector efficiency (0.57), ICF (1.65) and kf (3.30) were used. The error bars shown in 1D concentration profiles were calculated as $E = (C_i(1 - C_\mathrm{i})/N)^{1/2}$, in which $C_i = N_i/N$, N_i represents the number of i solute ions/atoms and N represents the total number of counts with the given bin. The bin size is 0.2 nm.

The chemical compositions (Extended Data Table 1) of the as-built Ti samples were determined by inductively coupled plasma atomic emission spectroscopy and glow discharge mass spectrometry.

Measurements of tensile mechanical properties

Extraction of tensile specimens and microstructural uniformity

DED-fabricated (Fig. 1a) and as-cast rectangular coupons were machined into dog-bone-like tensile specimens with gauge dimensions of 12 × 3 × 2 mm^3 (as per Australian Standard AS 1391-2007). Each tensile specimen was extracted from the middle nine layers along the length direction of the rectangular coupon. Supplementary Fig. 5 shows an array of ten as-fabricated coupons and the schematic extraction of each tensile sample. This is to ensure that each batch of tensile specimens has consistent microstructures for high repeatability. As shown in Supplementary Fig. 4, in the selected processing window (the green zone), stabilization of the average temperature occurs from the 8th or 9th layer of deposition. Although the

average temperature increases with increasing build height, the increase across every eight layers after layer 9 is limited to approximately 30 °C (negligible for microstructural evolution). Microstructural examination of different batches of tensile specimens along the build height confirmed consistent microstructures (see Supplementary Note 2).

Tensile testing

All tensile specimens were uniaxially tested at room temperature with an initial strain rate of 1×10^{-3} s^{-1} using an MTS universal testing facility (MTS 810, 100 kN) equipped with a non-contact laser extensometer.

Repeatability of tensile stress–strain curves

Supplementary Note 7 has briefly analysed the repeatability issue of the tensile stress–strain curves produced in this work (Extended Data Fig. 2) in accordance with ASTM E8/E8M-21 ('Standard Test Methods for Tension Testing of Metallic Materials'). The tensile data produced in this work meet the expected repeatability requirements.

The CALPHAD method

The Pandat software and PanTi2021 database developed by CompuTherm[43] were used to calculate the equilibrium volume fraction of the β-phase and the equilibrium partition of the O and Fe atoms between the α-phase and the β-phase. The use of the CALPHAD method in this work was restricted to the approximate isothermal holding stage, that is, at one of the predicted stabilized temperatures (such as 650 °C) in Fig. 1c. The predictions were used to help understand the microstructural evolution during this isothermal holding stage. It was not used to predict the far-from-equilibrium cooling process. The isothermal holding time at each predicted stabilized temperature is up to around 14 min in each build. It is calculated this way: the deposition time for each odd-number layer (for example, 35.85 s at 600 mm min^{-1}) + the deposition time for each even-number layer (for example, 37.71 s at 600 mm min^{-1}) + each layer-to-layer interval time (15 s).

Therefore, it is for isothermal approximation, although the deposition process is a far-from-equilibrium process.

DFT calculations

DFT calculations were performed using the generalized gradient approximation[44] and the projector augmented-wave method implemented in the VASP code[45]. The cutoff energy for plane-wave basis sets was 500 eV. Monkhorst-Pack k-points were meshed by 6 × 6 × 6, 8 × 8 × 8 and 6 × 6 × 1 for the 96-atom (and 128-atom), 54-atom bulk models and the interface model, respectively. For all systems, atomic relaxation was allowed until the forces were less than 0.01 eV $Å^{-1}$.

To build an α/β (HCP/BCC) interface model, we first constructed two surface slab models. For this specific interface, the in-plane lattice constants of the BCC ($\bar{1}1\bar{2}$) surface slab are a = 4.6000 Å and b = 5.7260 Å, whereas the corresponding values for the HCP ($\bar{1}100$) surface slab are a = 4.6250 Å and b = 2.9236 Å. The a values are very similar but the BCC b value is approximately twice the HCP b value. Therefore, we constructed a 1 × 1 in-plane ten-layer BCC ($\bar{1}1\bar{2}$)/1 × 2 in-plane ten-layer HCP ($\bar{1}100$) interface structure (Fig. 4a, ten layers on each side) to simulate an α/β interface.

Data availability

All core data generated and analysed for this study can be found in this article, its Supplementary Information file, which contains 20 supplementary figures and 8 supplementary tables, and the source data spreadsheet files provided (Fig. 1b,c, Fig. 4, Extended Data Fig. 5, Supplementary Fig. 4 and Supplementary Table 8). Further source data (>40 GB) leading to these core data are available from the corresponding authors without any restrictions.

References

1. Moiseyev, V. N. *Titanium Alloys: Russian Aircraft and Aerospace Applications* (CRC Press, 2005).

2. Donachie, M. J. *Titanium: A Technical Guide* 2nd edn (ASM International, 2000).

3. Welsch, G., Boyer, R. & Collings, E. *Materials Properties Handbook: Titanium Alloys* (ASM International, 1993).

4. Lütjering, G. & Williams, J. C. *Titanium* 2nd edn (Springer, 2007).

5. Polmear, I., StJohn, D., Nie, J.-F. & Qian, M. *Light Alloys: Metallurgy of the Light Metals* 5th edn (Butterworth-Heinemann, 2017).

6. Yu, Q. et al. Origin of dramatic oxygen solute strengthening effect in titanium. *Science* **347**, 635–639 (2015).

7. Barkia, B., Couzinié, J.-P., Lartigue-Korinek, S., Guillot, I. & Doquet, V. In situ TEM observations of dislocation dynamics in α titanium: effect of the oxygen content. *Mater. Sci. Eng. A* **703**, 331–339 (2017).

8. University of California - Berkeley. How oxygen is like kryptonite to titanium. *ScienceDaily* https://www.sciencedaily.com/releases/2015/02/150205142919.htm (2015)

9. Zeng, W. & Zhou, Y. Effect of beta flecks on mechanical properties of Ti–10V–2Fe–3Al alloy. *Mater. Sci. Eng. A* **260**, 203–211 (1999).

10. Marui, Y., Kinoshita, T. & Takahashi, K. Development of a titanium material by utilizing off-grade titanium sponge. *SAE Tech. Pap.* **32**, 1816 (2002).

11. Takeda, O., Ouchi, T. & Okabe, T. H. Recent progress in titanium extraction and recycling. *Metall. Mater. Trans. B* **51**, 1315–1328 (2020).

12. Gao, F. et al. Environmental impacts analysis of titanium sponge production using Kroll process in China. *J. Clean. Prod.* **174**, 771–779 (2018).

13. Bermingham, M. J., McDonald, S. D., StJohn, D. H. & Dargusch, M. S. Beryllium as a grain refiner in titanium alloys. *J. Alloys Compd.* **481**, L20–L23 (2009).

14. Lim, J., McMahon, C., Pope, D. & Williams, J. The effect of oxygen on the structure and mechanical behavior of aged Ti-8 Wt Pct Al. *Metall. Trans. A* **7**, 139–144 (1976).

15. Vavilova, V. V. Effect of oxygen on the properties of titanium and its alloys. *Met. Sci. Heat Treat.* **15**, 838–841 (1973).

16. ATI 425® Alloy, (Grade 38). Technical Data Sheet. http://atimaterials.com/Products/Documents/datasheets/titanium/alloyed/ati_425_alloy_tds_en_v5.pdf (Allegheny Technologies Incorporated, 2013).

17. Bania, P. J. Beta titanium alloys and their role in the titanium industry. *JOM* **46**, 16–19 (1994).

18. Mehjabeen, A., Xu, W., Qiu, D. & Qian, M. Redefining the β-phase stability in Ti-Nb-Zr alloys for alloy design and microstructural prediction. *JOM* **70**, 2254–2259 (2018).

19. Fujii, H. Development of low-cost high-strength Ti-Fe-O-N alloy series. *Nippon Steel Tech. Rep.* **85**, 107–112 (2002).

20. Koike, M., Ohkubo, C., Sato, H., Fujii, H. & Okabe, T. Evaluation of cast Ti–Fe–O–N alloys for dental applications. *Mater. Sci. Eng. C* **25**, 349–356 (2005).

21. Yin, L. & Umezawa, O. Influence of dispersed β phase on local deformation pattern in mill-annealed Ti-Fe-O alloy under dwell fatigue loading. *Int. J. Fatigue* **152**, 106447 (2021).

22. Yokoyama, H., Umezawa, O., Nagai, K., Suzuki, T. & Kokubo, K. Cyclic deformation, dislocation structure, and internal fatigue crack generation in a Ti-Fe-O alloy at liquid nitrogen temperature. *Metall. Mater. Trans. A* **31**, 2793–2805 (2000).

23. Takahashi, K. & Sato, E. Mechanical properties of Ti-6Al-4V alloy using off-grade sponge titanium as raw material. *J. Iron Steel Inst. Jpn.* **98**, 33–38 (2012).

24. Simonelli, M. et al. The influence of iron in minimizing the microstructural anisotropy of Ti-6Al-4V produced by laser powder-bed fusion. *Metall. Mater. Trans. A* **51**, 2444–2459 (2020).

25. Song, T. et al. Simulation-informed laser metal powder deposition of Ti-6Al-4V with ultrafine α-β lamellar structures for desired tensile properties. *Addit. Manuf.* **46**, 102139 (2021).

26. Carroll, B. E., Palmer, T. A. & Beese, A. M. Anisotropic tensile behavior of Ti–6Al–4V components fabricated with directed energy deposition additive manufacturing. *Acta Mater.* **87**, 309–320 (2015).

27. Todaro, C. J. et al. Grain structure control during metal 3D printing by high-intensity ultrasound. *Nat. Commun.* **11**, 142 (2020).

28. Zhao, Z. et al. Formation mechanism of the α variant and its influence on the tensile properties of laser solid formed Ti-6Al-4V titanium alloy. *Mater. Sci. Eng. A* **691**, 16–24 (2017).

29. Luo, S. D. et al. High oxygen-content titanium and titanium alloys made from powder. *J. Alloys Compd.* **836**, 155526 (2020).

30. Xia, Y., Zhao, J., Tian, Q. & Guo, X. Review of the effect of oxygen on titanium and deoxygenation technologies for recycling of titanium metal. *JOM* **71**, 3209–3220 (2019).

31. Lazić, I., Bosch, E. G. & Lazar, S. Phase contrast STEM for thin samples: integrated differential phase contrast. *Ultramicroscopy* **160**, 265–280 (2016).

32. Cottrell, A. H. & Bilby, B. A. Dislocation theory of yielding and strain ageing of iron. *Proc. Phys. Soc. A* **62**, 49–62 (1949).

33. Wang, S., Aindow, M. & Starink, M. Effect of self-accommodation on α/α boundary populations in pure titanium. *Acta Mater.* **51**, 2485–2503

(2003).

34. Watanabe, T. An approach to grain boundary design for strong and ductile polycrystals. *Res Mech.* **11**, 47–84 (1984).

35. Lu, K., Lu, L. & Suresh, S. Strengthening materials by engineering coherent internal boundaries at the nanoscale. *Science* **324**, 349–352 (2009).

36. Hull, D. & Bacon, D. J. *Introduction to Dislocations* 245–247 (Elsevier, 2011).

37. Anderson, P. M., Hirth, J. P. & Lothe, J. *Theory of Dislocations* 3rd edn 603–604 (Cambridge Univ. Press, 2017).

38. Simulating Welding With Simufact Welding. *Hexagon* https://www.simufact.com/simufactwelding-welding-simulation.html (2022).

39. Michaleris, P. Modeling metal deposition in heat transfer analyses of additive manufacturing processes. *Finite Elem. Anal. Des.* **86**, 51–60 (2014).

40. Sun, Y. & Hao, M. Statistical analysis and optimization of process parameters in Ti6Al4V laser cladding using Nd:YAG laser. *Opt. Lasers Eng.* **50**, 985–995 (2012).

41. Xu, W. et al. Additive manufacturing of strong and ductile Ti–6Al–4V by selective laser melting via in situ martensite decomposition. *Acta Mater.* **85**, 74–84 (2015).

42. Cao, S. et al. Role of martensite decomposition in tensile properties of selective laser melted Ti-6Al-4V. *J. Alloys Compd.* **744**, 357–363 (2018).

43. Cao, W. et al. PANDAT software with PanEngine, PanOptimizer and PanPrecipitation for multi-component phase diagram calculation and materials property simulation. *Calphad* **33**, 328–342 (2009).

44. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

45. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758 (1999).

Acknowledgements

We gratefully acknowledge support from the Australian Research Council (DP180103205, DP220103407, DP200100940, DP200102666, DP190102243 and IC180100005), the Australia–US Multidisciplinary University Research Initiative programme supported by the Australian Government through the Department of Defence under the Next Generation Technologies Fund, the Research Committee of The Hong Kong Polytechnic University (PolyU) (Project code: CD4F and UAMT), PolyU Research and Innovation Office (Project code: BBR5 and BBX2), and the funding support for the State Key Laboratories in Hong Kong from the Innovation and Technology Commission of the Government of the Hong Kong Special Administrative Region, China. We thank the RMIT Advanced Manufacturing Precinct (AMP), RMIT Microscopy & Microanalysis Facility (especially M. Field) and Sydney Microscopy & Microanalysis at the University of Sydney, which is a node of Microscopy Australia, for their facilities. We thank other team members for their contributions, including S. Luo for discussions on adding oxygen to titanium, A. Jones for printing the samples for Supplementary Fig. 15 (assisted by Z. Wu) and Supplementary Fig. 19 (assisted by Q. Zhou), Q. Zhou for performing the tensile tests for Supplementary Fig. 19, the technical team at RMIT AMP for machining all samples of this work from January 2019 to February 2023 and R. Hu for the porosity analysis of the as-cast samples. T.S. thanks A. Jones for training her on the TRUMPF TruLaser Cell 7020 system. Our DFT calculations were supported by the National Computational Infrastructure (NCI), with expert facilitation by the Sydney Informatics Hub team at the University of Sydney. Both Microscopy Australia and the NCI are supported by the Australian Government's National Collaborative Research Infrastructure Scheme.

Author information

Author notes

1. These authors contributed equally: Tingting Song, Zibin Chen

Authors and Affiliations

1. Centre for Additive Manufacturing, School of Engineering, RMIT University, Melbourne, Victoria, Australia

   Tingting Song, Shenglu Lu, Milan Brandt & Ma Qian

2. School of Aerospace, Mechanical and Mechatronic Engineering, The University of Sydney, Sydney, New South Wales, Australia

   Zibin Chen, Xiangyuan Cui, Hansheng Chen, Hao Wang, Xiaozhou Liao & Simon P. Ringer

3. Australian Centre for Microscopy & Microanalysis, The University of Sydney, Sydney, New South Wales, Australia

   Zibin Chen, Xiangyuan Cui, Hansheng Chen, Hao Wang, Xiaozhou Liao & Simon P. Ringer

4. Research Institute for Advanced Manufacturing, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China

   Zibin Chen, Bailiang Qin & Kang Cheung Chan

5. State Key Laboratory of Ultra-precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China

   Zibin Chen & Kang Cheung Chan

6. Hexagon Manufacturing Intelligence, Doncaster, Victoria, Australia

   Tony Dong

Contributions

M.Q. and T.S. designed the experimental alloys. T.D. and T.S. performed simulations. T.S., M.Q. and M.B. conceived the AM process. T.S. printed the samples and completed essential characterization, data analysis and compilation. Z.C., X.L. and S.P.R. led the atomic-scale characterization, data analyses, DFT calculations and proposed the strengthening mechanisms. Z.C. conducted electron microscopy experiments and analyses. X.C. carried out DFT calculations. H.C. conducted APT experiments. S.L. performed EBSD analyses. M.Q. and S.L. analysed fracture behaviours. H.W. assisted in electron microscopy experiments and EBSD analyses. B.Q. and K.C.C. manufactured ingot samples. M.Q., M.B., S.P.R., X.L. and Z.C. acquired funding. M.Q. and T.S. led the response to reviews. All authors critically reviewed the results and edited the manuscript drafted by the corresponding authors.

Corresponding authors

Correspondence to Simon P. Ringer or Ma Qian.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Marco Simonelli and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Microstructure of the as-cast reference alloy Ti–0.35O–3Fe (measured: Ti–0.33O–3.11Fe).

a, An overview of the microstructure comprising α–β lamellae and β-flecks. **b**, A closer view of a selected β-fleck. **c**, Energy-dispersive spectrometry (EDS) mapping of Fe in **b** to confirm the Fe-enriched β-fleck. **d**, A selected β-fleck phase for EDS line-scan analysis. Also shown is the as-cast porosity. **e**, EDS line scan further confirms the high Fe content in the β-fleck shown in **d**. **f**, A backscattered electron image showing the α–β lamellae in this as-cast alloy.

Extended Data Fig. 2 Engineering stress–strain curves for each DED-fabricated Ti–O–Fe alloy together with the reference Ti–6Al–4V alloy and ultralow-Fe Ti–0.16O–0.062Fe alloy.

a, Changing alloy composition but not the DED conditions (scan speed: 800 mm min^{-1}, layer interval: 15 s). **b**, Control studies to vary the microstructure through changing the DED conditions but not the alloy composition (Ti–0.35O–3Fe). **c**, Ti–0.35O–3Fe (by design) fabricated by water-cooled copper-mould casting as part of the control studies.

Extended Data Fig. 3 Control studies to vary the microstructure of the Ti–0.35O–3Fe alloy by varying the manufacturing conditions as specified in Extended Data Table 2.

a–**e** By DED with different scan speeds and layer interval times. **f**, By water-cooled copper-mould casting (see Methods).

Extended Data Fig. 4 APT 3D reconstruction data.

a, Distribution of Ti, O and Fe in an α–β phase region in the DED-fabricated Ti–0.14O–3.23Fe alloy (800 mm min^{-1}, 15 s). The Fe-enriched blue phase is the β-phase, whereas the surrounding Fe-depleted phase is the α-phase. The thresholds for the isocomposition surface for O and Fe are 3 at% and 15 at%, respectively. The violet cylindrical bar represents the 45-nm-long range for the APT profiles shown in the right column, whereas the arrow represents the direction. The Fe-enriched β-phase contains no detectable O by APT. **b**,**c**, Distribution of Ti, Al, V, Fe and O in an α–β region in DED-fabricated Ti–6Al–4V–0.22O–0.20Fe alloy (800 mm min^{-1}, 15 s). The two black crack-like planes are isoconcentration surfaces of 10 at% V, representing the α–β phase boundaries. The downward vertical arrow in **b** indicates the direction of the 25-nm-long range for the APT profiles shown in **c**. The Fe-enriched and V-enriched region (between the two black crack-like planes) is the β-phase, which lies between two α-phases. Refer to Supplementary Table 4 for the quantitative results.

Extended Data Fig. 5 EBSD analyses of the crystallographic arrangements of the α-variants in DED-fabricated Ti–O–Fe alloys and reference alloy Ti–6Al–4V–0.22O–0.20Fe.

More than 10^5 α-variants were analysed in each alloy using EBSD. When considering only abutting α-variant lamellae (that is, by removing the intermediary β-phase lath), the misorientation \([11\bar{2}0]\)/60° is prevalent in the Ti–O–Fe alloys, whereas the misorientation \([\bar{10}55\bar{3}]\)/63.26°) is prevalent in the Ti–6Al–4V–0.22O–0.20Fe alloy. Alloys **a**–**d** and **f** were fabricated with the same DED conditions (800 mm min^{-1}, 15 s), whereas alloy **e** was fabricated with a zero-layer interval, that is, 800 mm min^{-1}, 0 s. The other conditions were unchanged (Extended Data Table 2).

Source data

Extended Data Fig. 6 TEM observations of dislocations in the α and β phases of the Ti–0.34O–3.25Fe alloy after tensile testing (ε_f = 9.0 ± 0.8%, σ_{UTS} = 1,157 ± 3 MPa).

The area selected is close to the tensile fracture surface. **a**,**b**,**d**,**e**, Bright-field images, in which **b** and **e** are closer views. **c**,**f**, Dark-field image. Dislocations are best seen in **c** and **f**. As expected, profound dislocations were observed in the β-phase, which is virtually free of O, whereas dislocations were also observed in the α-phase, which was virtually free of Fe.

Extended Data Fig. 7 TEM observations of dislocations in the α and β phases of the Ti–0.67O–3.30Fe alloy after tensile testing ($\varepsilon = 3.0 \pm 0.8\%$, $\sigma_{UTS} = 1,271 \pm 6$ MPa).

The area selected is close to the tensile fracture surface. **a**,**b**,**d**,**e**, Bright-field images, in which **b** and **e** are closer views. **c**,**f**, Dark-field image. Dislocations are best seen in **c** and **f**. As expected, more dislocations were observed in the β-phase, which is virtually free of O, whereas few dislocations were observed in the α-phase owing to its high O content.

Extended Data Fig. 8 Tensile fractographs of each DED-fabricated Ti–O–Fe alloy (800 mm min^{-1}, 15 s).

The tensile properties of each alloy are listed in Extended Data Table 1.

Extended Data Table 1 Composition and tensile properties of as-fabricated alloys
Extended Data Table 2 Processing parameters and scan path for simulation and fabrication

Supplementary information

Supplementary Information

This file contains Supplementary Figs. 1–20, Supplementary Tables 1–8, Supplementary Notes 1–7 and Supplementary References.

Peer Review File

Supplementary Fig. 4 source data

Supplementary Table 8 source data

Source data

Source Data Fig. 1

Source Data Fig. 4

Source Data Extended Data Fig. 5

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Song, T., Chen, Z., Cui, X. *et al.* Strong and ductile titanium–oxygen–iron alloys by additive manufacturing. *Nature* **618**, 63–68 (2023).

https://doi.org/10.1038/s41586-023-05952-6

- Received: 08 November 2021
- Accepted: 14 March 2023
- Published: 31 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05952-6

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

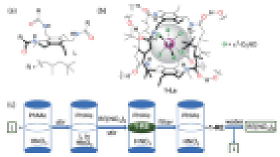
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Designer titanium alloys created using 3D printing

Nature Research Briefing 31 May 2023

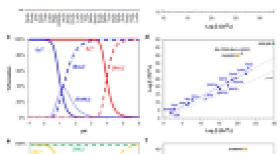
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05952-6



Selective separation of light rare-earth elements by supramolecular encapsulation and precipitation

03 August 2022

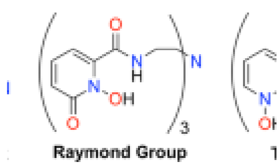
Joseph G. O’Connell-Danes, Bryne T. Ngwenya, … Jason B. Love



Ultra-selective ligand-driven separation of strategic actinides

04 June 2019

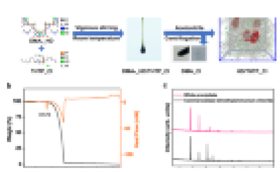
Gauthier J.-P. Deblonde, Abel Ricano & Rebecca J. Abergel



High-throughput screening for discovery of benchtop separations systems for selected rare earth elements

20 January 2020

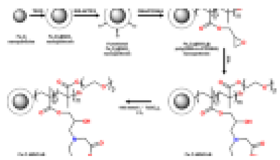
Joshua J. M. Nelson, Thibault Cheisson, … Eric J. Schelter



[Type II porous ionic liquid based on metal-organic cages that enables l-tryptophan identification]

29 April 2022

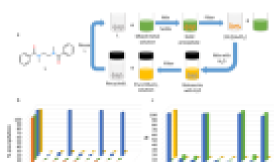
Zhuxiu Zhang, Baolin Yang, … Xu Qiao



[Effect of shell structure of Ti-immobilized metal ion affinity chromatography core-shell magnetic particles for phosphopeptide enrichment]

31 October 2019

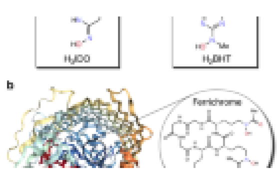
Anna Laura Capriotti, Michela Antonelli, … Aldo Laganà



[Tuneable separation of gold by selective precipitation using a simple and recyclable diamide]

29 October 2021

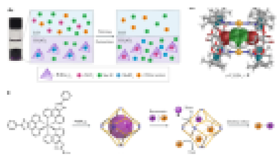
Luke M. M. Kinsman, Bryne T. Ngwenya, … Jason B. Love



[Siderophore-inspired chelator hijacks uranium from aqueous medium]

18 February 2019

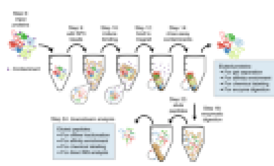
Alexander S. Ivanov, Bernard F. Parker, … Ilja Popovs



Metal–organic cages for molecular separations

27 January 2021

Dawei Zhang, Tanya K. Ronson, … Jonathan R. Nitschke



Single-pot, solid-phase-enhanced sample preparation for proteomics experiments

21 November 2018

Christopher S. Hughes, Sophie Moggridge, … Jeroen Krijgsveld

- Article
- Open Access
- Published: 31 May 2023

Enhanced rare-earth separation with a metal-sensitive lanmodulin dimer

- Joseph A. Mattocks,

- Jonathan J. Jung,
- Chi-Yun Lin,
- Ziye Dong,
- Neela H. Yennawar,
- Emily R. Featherston,
- Christina S. Kang-Yun,
- Timothy A. Hamilton,
- Dan M. Park ,
- Amie K. Boal &
- …
- Joseph A. Cotruvo Jr

Show authors

Nature **volume 618**, pages 87–93 (2023)

- 2865 Accesses
- 109 Altmetric
- Metrics details

Abstract

Technologically critical rare-earth elements are notoriously difficult to separate, owing to their subtle differences in ionic radius and coordination number[1,2,3]. The natural lanthanide-binding protein lanmodulin (LanM)[4,5] is a sustainable alternative to conventional solvent-extraction-based separation[6]. Here we characterize a new LanM, from *Hansschlegelia quercus* (*Hans*-LanM), with an oligomeric state sensitive to rare-earth ionic radius, the lanthanum(III)-induced dimer being >100-fold tighter than the dysprosium(III)-induced dimer. X-ray crystal structures illustrate how picometre-scale differences in radius between lanthanum(III) and dysprosium(III) are propagated to *Hans*-LanM's quaternary structure through a carboxylate shift that rearranges a second-sphere hydrogen-bonding network. Comparison to the prototypal LanM from *Methylorubrum extorquens* reveals distinct metal coordination strategies, rationalizing *Hans*-

LanM's greater selectivity within the rare-earth elements. Finally, structure-guided mutagenesis of a key residue at the *Hans*-LanM dimer interface modulates dimerization in solution and enables single-stage, column-based separation of a neodymium(III)/dysprosium(III) mixture to >98% individual element purities. This work showcases the natural diversity of selective lanthanide recognition motifs, and it reveals rare-earth-sensitive dimerization as a biological principle by which to tune the performance of biomolecule-based separation processes.

Main

The irreplaceable roles of rare-earth (RE) elements in ubiquitous modern technologies ranging from permanent magnets to light-emitting diodes and phosphors have renewed interest in one of the grand challenges of separation science—efficient separation of lanthanides[1]. The separation of these 15 elements is complicated by the similar physicochemical properties of their predominating +III ions, with ionic radii decreasing only 0.19 Å between La^{III} and Lu^{III} (ref. [7]), which also leads to these metals co-occurring in RE-bearing minerals. Conventional hydrometallurgical liquid–liquid extraction methods for RE production utilize organic solvents such as kerosene and toxic phosphonate extractants and require dozens or even hundreds of stages to achieve high-purity individual RE oxides[3,8]. The inefficiency and large environmental impact of RE separations[9] have stimulated research efforts into alternative ligands with larger separation factors between adjacent REs[10,11,12,13,14], and greener process designs to achieve RE separation in fewer stages[15] and using all-aqueous chemistry[6,16,17,18,19,20].

The discovery of the founding member of the LanM family of lanthanide-binding proteins demonstrated that nature has evolved macromolecules surpassing the selectivity of synthetic f-element chelators[4]. The prototypal LanM, from *M. extorquens* AM1 (*Mex*-LanM), is a small (12-kDa), monomeric protein that undergoes a selective conformational response to picomolar concentrations of lanthanides[4,18] and actinides[21,22,23,24], has facilitated understanding of lanthanide uptake in methylotrophs[25], and has served as a technology platform for f-element detection[26], recovery[18,27] and separation[6]. Unusually among RE chelators, *Mex*-LanM favours the larger

and more abundant light REs (LREs), especially La^{III}–Sm^{III}, over heavy REs (HREs)[4]. Our recent demonstration that even single substitutions to the metal-binding motifs of *Mex*-LanM can improve actinide/lanthanide separations[23] spurred us to investigate whether orthologues of *Mex*-LanM might possess distinct, and potentially useful, metal selectivity trends.

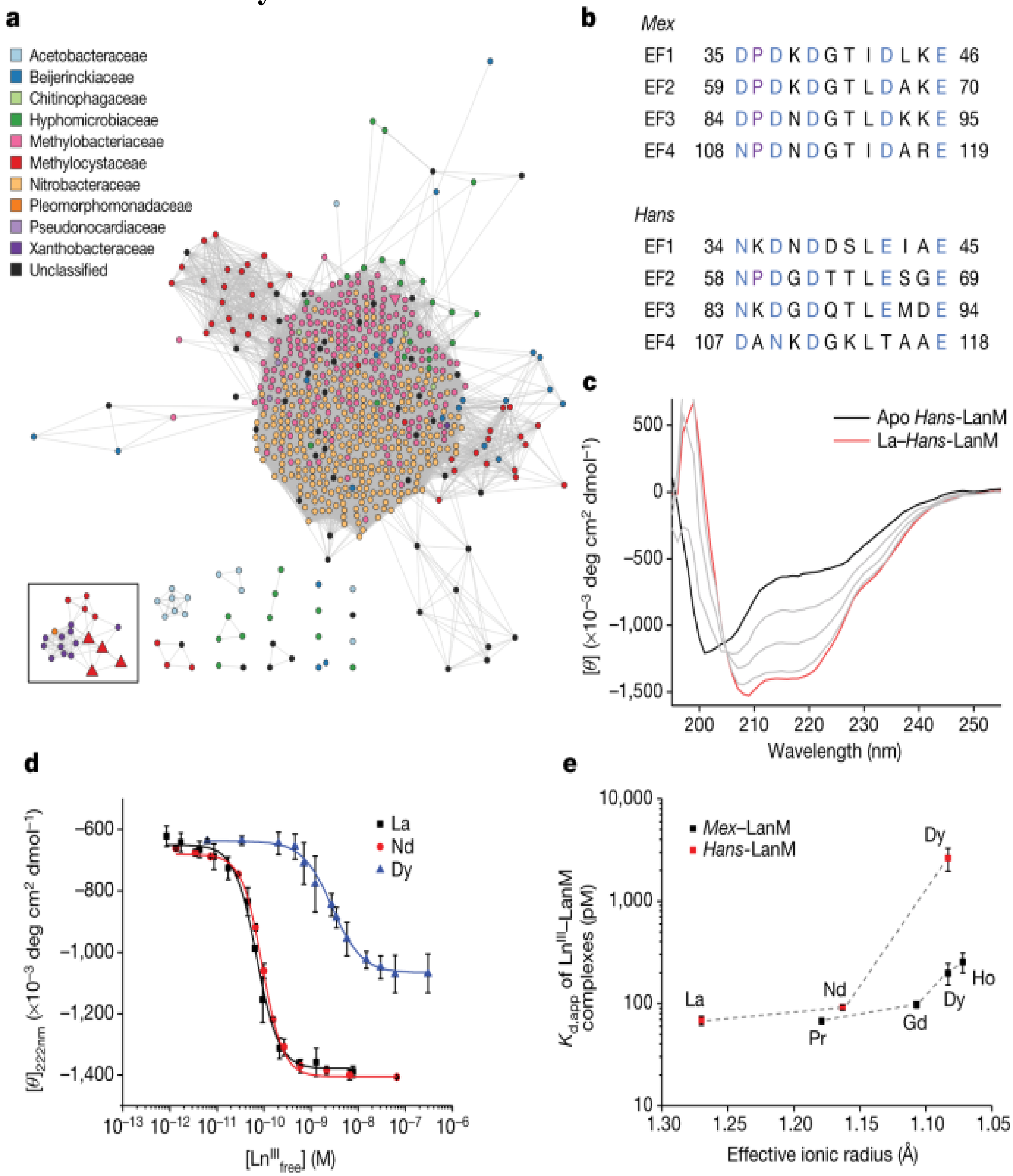
Herein, we report that the LanM from *Hansschlegelia quercus* (*Hans*-LanM), a methylotrophic bacterium isolated from English oak buds[28], exhibits enhanced RE separation capacity relative to *Mex*-LanM. Whereas *Mex*-LanM is always monomeric, *Hans*-LanM exists in a monomer/dimer equilibrium, the position of which depends on the specific RE bound. Three X-ray crystal structures of LanMs and structure-guided mutagenesis explain *Hans*-LanM's RE-dependent oligomeric state and its greater separation capacity than that of *Mex*-LanM. Finally, we leverage these findings to achieve single-stage *Hans*-LanM-based separation of the critical neodymium/dysprosium pair. These results illustrate how intermolecular interactions—common in proteins but rare in small molecules—may be exploited to improve RE separations.

Hans-LanM's distinct selectivity profile

We have proposed[4] several hallmarks of a LanM. First, LanMs possess four EF-hand motifs. EF hands comprise 12-residue, carboxylate-rich metal-binding loops flanked by α-helices, which traditionally respond to Ca^{II} binding;[29] in *Mex*-LanM, however, EF hands 1–3 bind lanthanide(III) ions with low-picomolar affinity and 10^8-fold selectivity over Ca^{II}, resulting in a large, lanthanide-selective disorder-to-order conformational transition[4]. EF4 binds with only micromolar affinity. Second, adjacent EF hands in LanMs are separated by 12–13 residues—rather than the typical ≈25 residues in Ca^{II}-responsive EF-hand proteins—resulting in an unusual three-helix bundle architecture with the metal-binding sites on the periphery[5]. Third, at least one EF hand contains proline at the second position (in *Mex*-LanM, all four EF hands feature P_2 residues). We searched sequence databases using the first two criteria and a sequence length of <200 residues, identifying 696 putative LanMs. These sequences were visualized using a sequence similarity network[30] to identify LanM sequences that cluster separately from

Mex-LanM. Notably, at a 65% identity threshold, a small cluster of sequences that is remote from the main cluster of 642 sequences is formed (Fig. 1a). This exclusive cluster (the *Hans* cluster), includes bacteria from several genera, including *Hansschlegelia* and *Xanthobacter* (Extended Data Fig. 1), all of which are facultative methylotrophs[31].

Fig. 1: *Hans*-LanM diverges from *Mex*-LanM in sequence and RE versus RE selectivity.



a, Sequence similarity network of core LanM sequences indicates that *Hans*-LanM forms a distinct cluster. The sequence similarity network includes 696 LanM sequences connected with 48,647 edges, thresholded at a BLAST *E* value of 1×10^{-5} and 65% sequence identity. The black box encloses nodes clustered with *Hans*-LanM. The LanM sequence associated with *Mex* (downtriangle) and four within *Hansschlegelia* (uptriangle) are enlarged compared to other nodes (circles). Colours of the nodes represent the family from which the sequences originate. **b**, Comparison of the sequences of the four EF hands of *Mex*- and *Hans*-LanMs. Residues canonically involved in metal binding in EF hands are in blue; Pro residues are in purple. **c**, Circular dichroism spectra from a representative titration of *Hans*-LanM with La^{III}, showing the metal-associated conformational response increasing helicity; apoprotein is bold black, La^{III}-saturated protein is bold red. **d**, Circular dichroism titration of *Hans*-LanM with La^{III}, Nd^{III} and Dy^{III} (pH 5.0). Each point represents the mean ± s.d. from three independent experiments. **e**, Comparison of $K_{d,app}$ values (pH 5.0) for *Mex*-LanM (black[18]) and *Hans*-LanM (red), plotted versus ionic radius[7]. Mean ± s.e.m. from three independent experiments.

Source Data

Hans-LanM features low (33%) sequence identity with *Mex*-LanM (Supplementary Fig. 1) and divergent EF-hand motifs, particularly at the first, second and ninth positions (Fig. 1b), which are important positions in *Mex*-LanM[23,26] and other EF-hand proteins[29]. Therefore, *Hans*-LanM presented an opportunity to determine features essential for selective lanthanide recognition in LanMs.

Hans-LanM was expressed in *Escherichia coli* as a 110 amino acid protein (Supplementary Fig. 1). La^{III} and Nd^{III} were selected as representative LREs and Dy^{III} was selected as a representative HRE for complexation studies. The protein binds about three equivalents of La^{III} and Nd^{III}, and slightly less Dy^{III}, by inductively coupled plasma mass spectrometry (Supplementary Table 1), as does *Mex*-LanM[4]. Also like *Mex*-LanM[4], *Hans*-LanM exhibits little helical content in the absence of metal, as judged by the circular dichroism signal at 222 nm (Fig. 1c). Unexpectedly, only two equivalents of La^{III} or Dy^{III} were sufficient to cause *Hans*-LanM's complete

conformational change (Supplementary Fig. 2), indicating that the third binding equivalent is weak and does not increase helicity.

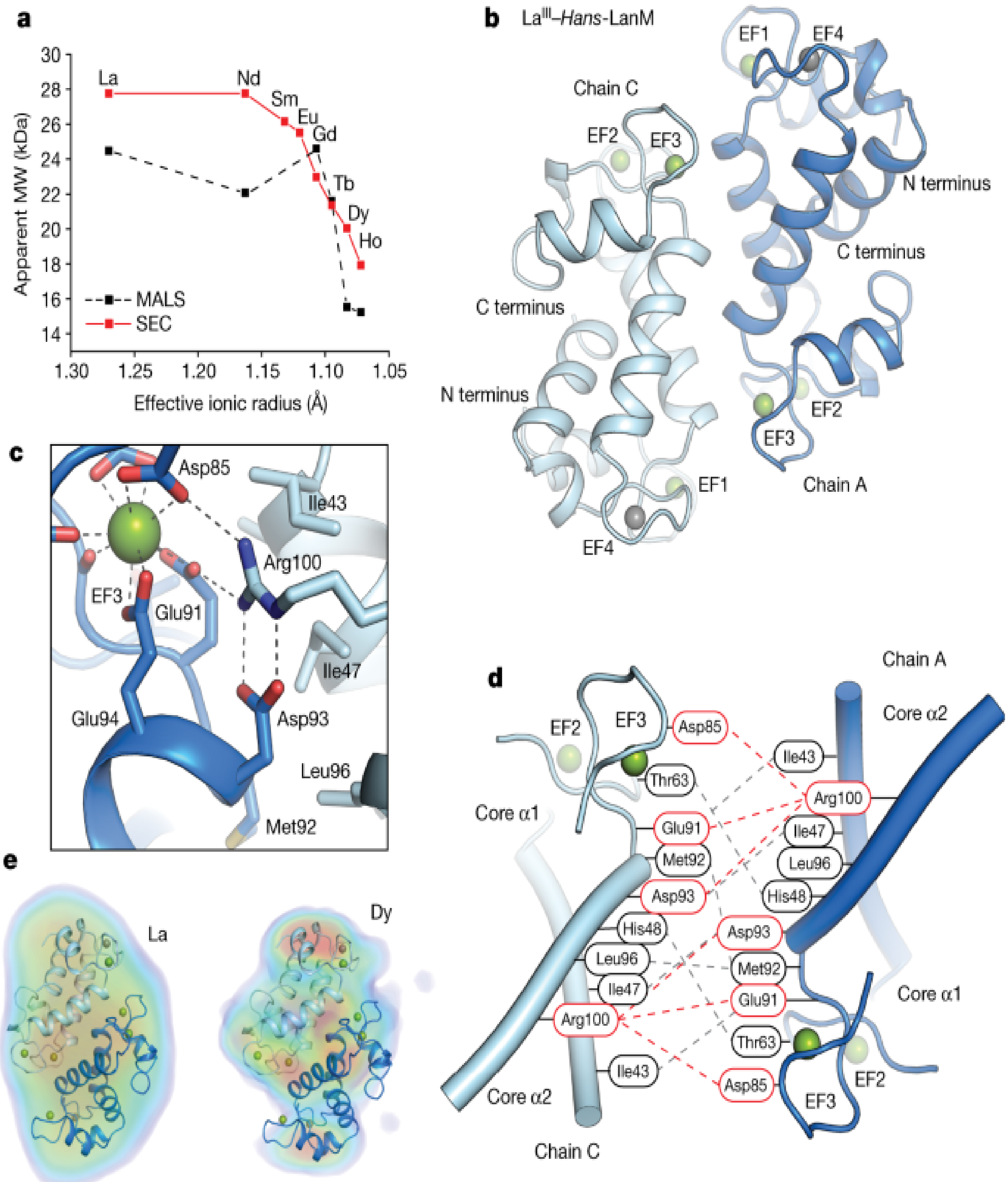
The apparent dissociation constants ($K_{d,app}$) determined by circular dichroism spectroscopy[4] reflect the RE versus RE, and RE versus non-RE, selectivities of *Mex*-LanM under competitive RE recovery conditions[6,18]. Therefore, similar determinations of $K_{d,app}$ with free metal concentrations controlled by a competitive chelator[4,32] were applied to *Hans*-LanM; the results (Fig. 1d and Supplementary Table 2) diverged from those for *Mex*-LanM. Binding of La^{III} and Nd^{III} to *Hans*-LanM increases molar ellipticity at 222 nm by 2.3-fold, the full conformational change evident in stoichiometric titrations. The conformational change is cooperative (Hill coefficients, n, of 2; Supplementary Table 2), and the $K_{d,app}$ values are similar, 68 and 91 pM, respectively. By contrast, even though Dy^{III} induces the same overall response as La^{III} in stoichiometric titrations (Supplementary Fig. 2), in the chelator-buffered Dy^{III} titrations *Hans*-LanM exhibits a lesser conformational response (1.8-fold increase). This difference indicates that at least one of the Dy^{III}-binding sites is very weakly responsive ($K_{d,app} > 0.3$ µM, the highest concentration accessible in the chelator-buffered titrations). The main response to Dy^{III} occurs at 2.6 nM, >30-fold higher than with the LREs, and with little or no cooperativity ($n = 1.3$). By contrast, *Mex*-LanM shows only a modest preference for LREs (about fivefold; Fig. 1e; ref. [4]), and all lanthanides and Y^{III} induce similar conformational changes and cooperativity[18]. *Hans*-LanM responds to calcium(II) weakly ($K_{d,app} = 60$ µM), with the same lack of cooperativity ($n = 1.0$) and partial conformational change evident with Dy^{III} (Extended Data Fig. 2). Therefore, *Hans*-LanM discriminates more strongly between LREs and HREs than does *Mex*-LanM, with the HRE complexes exhibiting lower affinity, lesser cooperativity and a lesser primary conformational change.

LRE-selective dimerization

The distinct behaviours of the LRE– and HRE–*Hans*-LanM complexes suggested mechanism(s) of LRE versus HRE selectivity not present in *Mex*-LanM. As *Mex*-LanM is a monomer in complex with LREs and HREs

alike[4,5], we considered that LREs and HREs might induce different oligomeric states in *Hans*-LanM. In the presence of three equivalents of La^{III}, *Hans*-LanM elutes from a size-exclusion chromatography (SEC) column not at the expected molecular weight (MW) of 11.9 kDa but instead at 27.8 kDa, suggestive of a dimer (Supplementary Figs. 3 and 4a). Starting gradually after Nd^{III} but sharply at Gd^{III}, the apparent MW decreases towards that expected for a monomer (Fig. 2a, Supplementary Fig. 4 and Supplementary Table 3). Notably, lanthanides heavier than Gd^{III} do not seem to support growth of RE-utilizing bacteria[33,34,35].

Fig. 2: A dimerization equilibrium sensitive to LRE versus HRE or non-RE coordination.



a, Apparent molecular weight of *Hans*-LanM complexes with REs as determined by analytical SEC (red lines) or SEC–MALS (black dashed line). See Supplementary Table 1 for conditions. Each individual data point is the result of a single experiment. **b**, The La^{III}-bound *Hans*-LanM dimer as determined by X-ray crystallography. La^{III} ions are shown as green spheres and Na^{I} ions are shown as grey spheres. **c**, Detailed view of the dimer

interface near EF3 of chain A (blue cartoon). Arg100 from chain C (light blue cartoon) anchors a hydrogen-bonding network involving Asp93 of chain A and two EF3 La^{III} ligands (Glu91 and Asp85). These interactions constitute the sole polar contacts at the dimer interface, providing a means to control the radius of the lanthanide-binding site at EF3. **d**, Schematic of the interactions at the dimer interface. Red dashed lines indicate hydrogen-bonding interactions and grey dashed lines indicate hydrophobic contacts. **e**, DENSS projections of electron density from small-angle X-ray scattering datasets for La^{III}-bound (left) and Dy^{III}-bound (right) *Hans*-LanM, overlaid with a PyMOL-generated ribbon diagram of the dimeric La^{III}–*Hans*-LanM crystal structure.

To provide further support for preferential dimerization in the presence of physiologically relevant LREs, RE complexes of *Hans*-LanM were analysed using multi-angle light scattering (MALS; Fig. 2a and Supplementary Fig. 5). The La^{III}, Nd^{III} and Gd^{III} complexes have MWs of 22–25 kDa, indicative of dimers, but MWs decrease starting with Tb^{III} and continue to Dy^{III} and Ho^{III}, at about 15 kDa (Extended Data Table 1), in agreement with the SEC data. Ca^{II}-bound *Hans*-LanM also indicated a MW of 14.7 kDa. The HRE–, Ca^{II}– and apo *Hans*-LanM complexes are still one-third larger than expected for a monomer, however, suggesting that these forms exist in a rapid equilibrium with ≈2:1 monomer/dimer ratio under these conditions. We next determined the K_d for dimerization (K_{dimer}) of apo, La^{III}-bound and Dy^{III}-bound *Hans*-LanM by isothermal titration calorimetry (Extended Data Table 2 and Supplementary Figs. 6–8). The apoprotein and Dy^{III}-bound protein weakly dimerize, with K_{dimer} values of 117 µM and 60 µM, respectively, consistent with the ratios of monomer and dimer reflected in the SEC and MALS traces. In the presence of La^{III}, however, the dimer was too tight to be able to observe monomerization by isothermal titration calorimetry, which indicates that K_{dimer} <0.4 µM (Supplementary Fig. 8). Thus, La^{III} favours *Hans*-LanM's dimerization by >100-fold over Dy^{III}.

A 1.8-Å-resolution X-ray crystal structure of *Hans*-LanM in complex with La^{III} confirms LRE-induced dimerization (Extended Data Fig. 3 and Supplementary Table 4). Two LanM monomers interact head-to-tail (Fig. 2b), burying about 600 Å^2 of surface area through hydrophobic and polar

contacts (Fig. 2c,d). These interactions occur largely between side chains contributed by the core helices α1 (between EF1 and EF2) and α2 (between EF3 and EF4; Supplementary Fig. 9). Residues at the dimer interface make direct contact with only one of the four metal-binding sites, EF3; three residues of EF3 in each monomer form a hydrogen-bonding network with Arg100 of the other monomer (Fig. 2c), suggesting that occupancy and coordination geometry at this site may control oligomeric state.

Hans-LanM and its complexes with three equivalents of La^{III}, Nd^{III} and Dy^{III} were also analysed by small-angle X-ray scattering (Supplementary Figs. 10 and 11). The calculated solvent envelopes[36] from the small-angle X-ray scattering data fit well to the crystallographic *Hans*-LanM dimer for La^{III}–*Hans*-LanM, adequately for Nd^{III}–*Hans*-LanM, but poorly for Dy^{III}–*Hans*-LanM (Fig. 2e and Supplementary Figs. 12–14). The weaker dimerization of Dy^{III}–*Hans*-LanM is also supported by quantitative metrics, such as the Porod volume (Supplementary Figs. 15 and 16 and Supplementary Tables 5 and 6). Together, the biochemical and structural results indicate that *Hans*-LanM's dimerization equilibrium depends strongly on the particular RE bound.

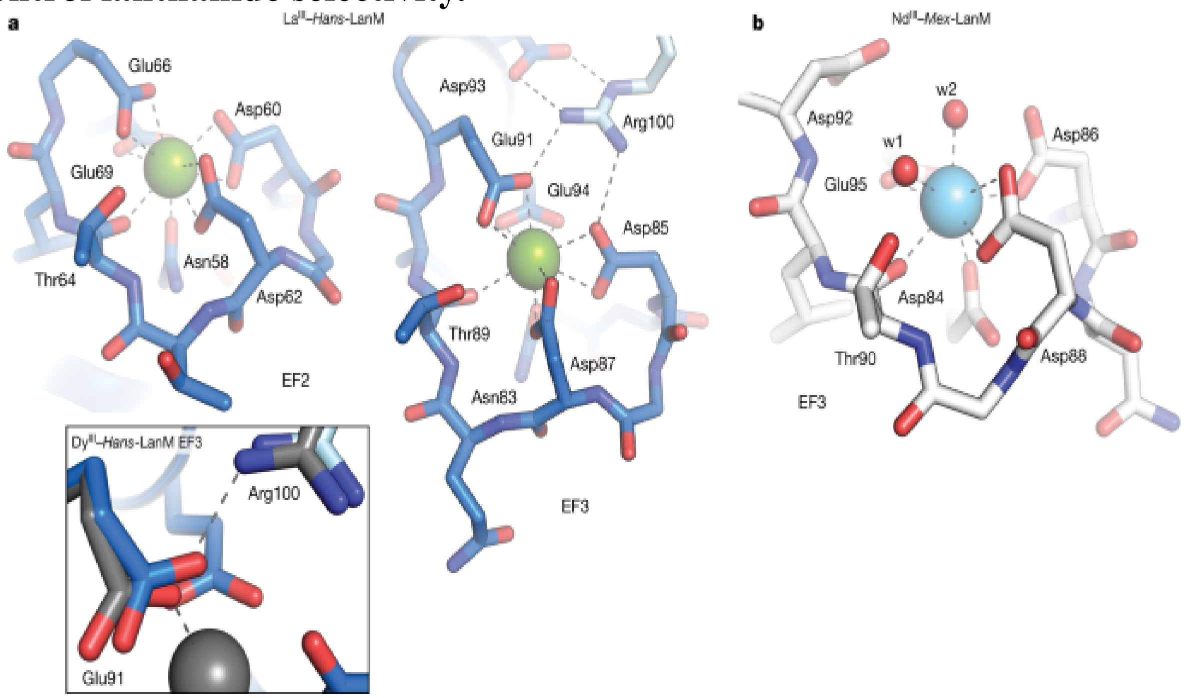
Structural basis for dimerization

The structure of La^{III}–*Hans*-LanM also provides one of the first detailed views of the coordination environments in a LanM, and indeed any natural biomolecule tasked with reversible lanthanide recognition. The previous NMR structure of *Mex*-LanM[5] revealed the protein's unusual fold, but it could not provide molecular-level details about the metal-binding sites. To understand the basis for LRE versus HRE discrimination, we also determined a 1.4-Å-resolution structure of Dy^{III}–*Hans*-LanM. Finally, we report a 1.01-Å-resolution structure of Nd^{III}–*Mex*-LanM, which rationalizes *Mex*-LanM's shallower RE selectivity trend[4].

In La^{III}–*Hans*-LanM, EF1–3 are occupied by La^{III} ions (Extended Data Fig. 3b–e). EF4 is structurally distinct, does not exhibit anomalous difference density consistent with La^{III} and was modelled with Na^{I} (Supplementary Fig. 17a). Each La^{III}-binding site is ten-coordinate, as observed in structures

of lanthanide-dependent methanol dehydrogenases[33,37] (Supplementary Fig. 18). A monodentate Asn (N_1 position), four bidentate Asp or Glu residues (D_3, D_5, E_9 and E_{12}) and a backbone carbonyl (T_7 or S_7) complete the first coordination sphere in EF1–3 (Fig. 3a). Exogenous solvent ligands are not observed (Supplementary Fig. 17b); luminescence studies of Eu^{III}–*Hans*-LanM to determine the number of coordinated solvent molecules (q) yielded $q = 0.11$, consistent with the absence of solvent ligands in the X-ray structure (Supplementary Fig. 19).

Fig. 3: *Hans*-LanM uses an extended hydrogen-bonding network to control lanthanide selectivity.



a, Zoomed-in views of EF2 (left) and EF3 (right) in La^{III}–*Hans*-LanM. La^{III} ions are shown as green spheres. Coordination bonds and hydrogen bonds are shown as dashed lines. Residues contributed by chain A are shown in blue and those contributed by chain C (in the case of EF3) are shown in light blue. Inset: overlay of La^{III}–*Hans*-LanM (blue and light blue) with Dy^{III}–*Hans*-LanM (grey), showing the carboxylate shift of Glu91 from bidentate (La) to monodentate (Dy). Coordination and hydrogen bonds (dashed lines) are shown only for the Dy case. **b**, Representative metal-binding site (EF3) in Nd^{III}–*Mex*-LanM. Nd^{III} ion is shown as an aqua sphere. Solvent molecules are shown as red spheres.

The lanthanide-binding sites in *Hans*-LanM additionally share extensive second-sphere interactions that may further constrain the positions of the ligands and the size of the metal-binding pore (Supplementary Fig. 20). This phenomenon is most obvious in EF3, at which the dimer interface mediates an extended hydrogen-bonding network involving several ligands. Arg100, contributed by the adjacent monomer, projects into the solvent-exposed side of EF3 to contact two carboxylate ligands, Asp85 (D_3) and Glu91 (E_9), enforcing their bidentate binding modes. Arg100 is also buttressed by Asp93 (EF3 D_{11}), unique to EF3 within *Hans*-LanM and not observed in *Mex*-LanM. We tested the importance of this network in *Hans*-LanM dimerization by making the minimal substitution, R100K. *Hans*-LanM(R100K) had nearly identical $K_{d,app}$ values and response to Nd^{III} and Dy^{III} as wild-type *Hans*-LanM, but the $K_{d,app}$ for La^{III} was twofold weaker (Supplementary Fig. 21 and Supplementary Table 7). SEC–MALS analysis indicated MWs of 10–13 kDa for apo, La^{III}– and Dy^{III}–*Hans*-LanM(R100K) (Supplementary Fig. 22 and Supplementary Table 8), indicative of increased monomerization, especially for the La^{III} complex, and suggesting that weaker dimerization may be responsible for the lower La^{III} affinity. All four residues comprising the Arg100–EF3 network are completely conserved in the *Hans* cluster (Supplementary Fig. 23), suggesting that these interactions may contribute to dimerization in these LanMs.

The structure of Dy^{III}–*Hans*-LanM confirms the importance of second-sphere control of ligand positioning (Extended Data Fig. 4, Supplementary Figs. 24–26 and Supplementary Tables 9 and 10). The overall structure of Dy^{III}–*Hans*-LanM is largely superimposable with that of La^{III}–*Hans*-LanM, and the coordination spheres of the Dy^{III} ions in EF1–3 are similar to those in La^{III}–*Hans*-LanM (Fig. 3a, inset), with the notable exception of E_9 (for example, Glu91 in EF3). This residue shifts from bidentate with La^{III} to monodentate with the smaller Dy^{III} ions, yielding a nine-coordinate distorted capped square antiprismatic geometry; the lower coordination number with a HRE ion is consistent with the case of other RE complexes[38,39]. In EF3, this carboxylate shift lengthens the distance between Arg100 and the proximal Oε of Glu91 from 2.9 Å (in La^{III}–*Hans*-LanM) to 3.2 Å (Supplementary Fig.

27). The rearrangement of this second-sphere hydrogen-bonding network suggests a structural basis for RE-dependent differences in K_{dimer} values.

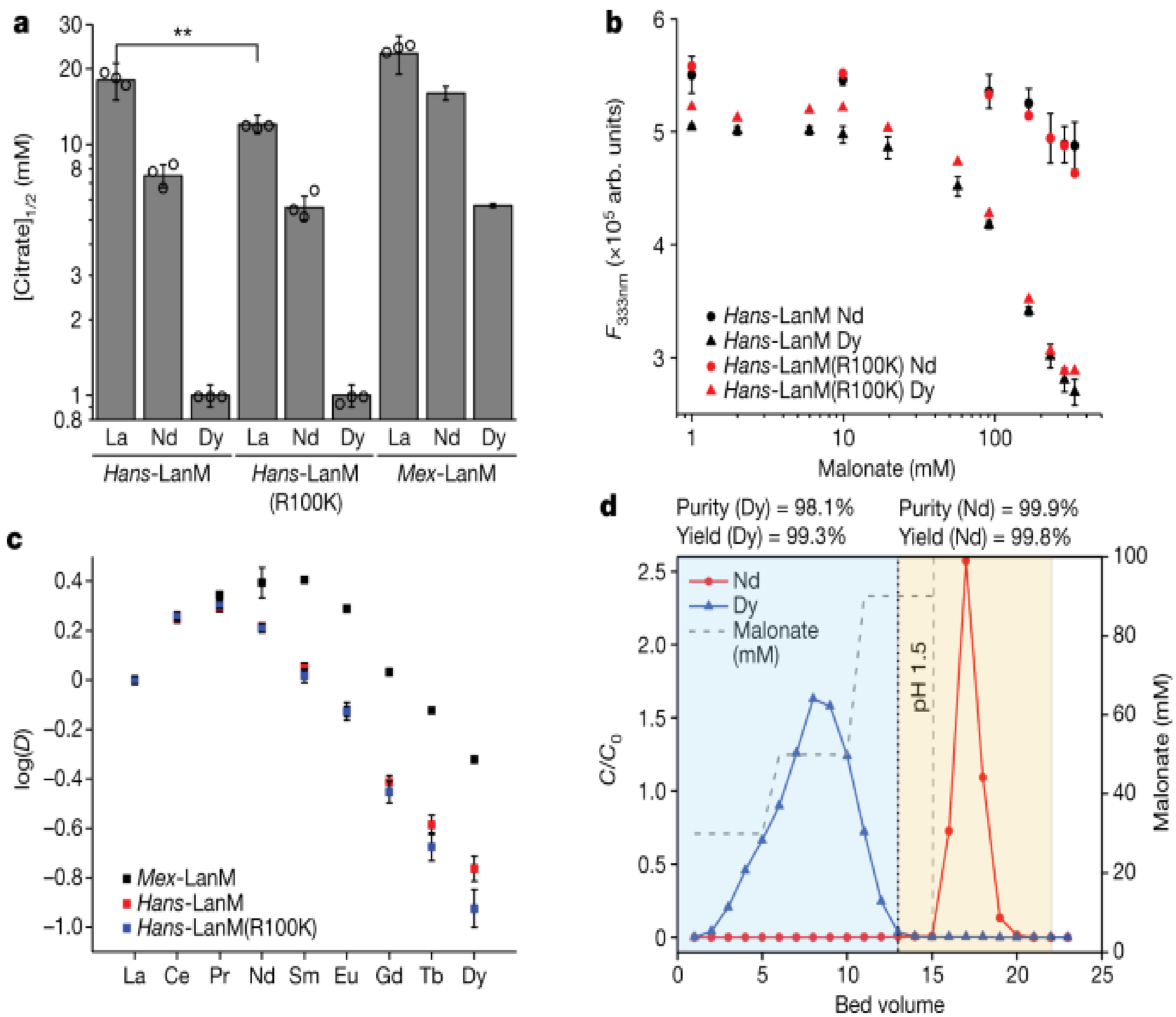
The metal-binding sites of *Mex*-LanM differ substantially from those of *Hans*-LanM. In *Mex*-LanM, all four EF hands are occupied by nine-coordinate (EF1–3) or ten-coordinate (EF4) Nd^{III} ions, each including two solvent ligands, not present in *Hans*-LanM (Fig. 3b and Supplementary Fig. 28). The observation of the two solvent molecules per metal site and the hydrogen bond to the D_9 residue validates recent spectroscopic studies[21,23,26]. The difference in coordination number between EF1–3 and EF4 is due to the D_3 carboxylate being monodentate in EF1–3 but bidentate in EF4. Although the Nd^{III} sites of *Mex*-LanM share the nine- and ten-coordination observed in Dy^{III}– and La^{III}–*Hans*-LanM, they structurally resemble the seven-coordinate Ca^{II}-binding sites of calmodulin (Supplementary Fig. 18). The increased coordination numbers in *Mex*-LanM relative to calmodulin result from bidentate coordination of D_5 and an additional solvent ligand. These similarities suggest that much of LanM's unique 10^8-fold selectivity for REs over Ca^{II} results from subtle differences in second-coordination-sphere and other more distal interactions. Finally, the exclusively protein-derived first coordination sphere in *Hans*-LanM, particularly due to coordination by E_9, yields more extended hydrogen-bonding networks (Supplementary Figs. 20 and 29) and probably enhances control over the radius of the binding site. Thus, the structures rationalize the extraordinary RE versus non-RE selectivity of *Mex*-LanM and *Hans*-LanM while also accounting for their differences in LRE versus HRE selectivity.

Single-stage Nd^{III}/Dy^{III} separation

The differences in stability and structure between *Hans*-LanM's LRE versus HRE complexes suggested that *Hans*-LanM (wild type and/or R100K) would outperform *Mex*-LanM in RE/RE separations. We focused on separating the RE pair of Nd^{III} and Dy^{III} used in permanent magnets. We first assayed the stabilities of the wild-type *Hans*-LanM and *Hans*-LanM(R100K) RE complexes against citrate, previously used as a desorbent with *Mex*-LanM[6]. RE–*Hans*-LanM complexes are generally less stable

against citrate than those of *Mex*-LanM, as expected on the basis of lower affinity (Fig. 1e), but the difference in stability between the Nd^{III}–*Hans*-LanM and Dy^{III}–*Hans*-LanM complexes—expressed as the ratio of citrate concentration required for 50% desorption of each metal ($[\text{citrate}]_{1/2}$), as reported by the fluorescence of *Hans*-LanM's two Trp residues (Supplementary Fig. 30)—is twofold greater than for *Mex*-LanM complexes (Fig. 4a, Supplementary Table 11 and Extended Data Fig. 5). Furthermore, the R100K substitution significantly destabilizes *Hans*-LanM's La^{III} complex against citrate, whereas it only slightly affects the Nd^{III} complex and does not affect the Dy^{III} complex. This result confirms that dimerization selectively stabilizes *Hans*-LanM's LRE complexes (and especially the La^{III} complex), a factor abrogated by the R100K substitution. Using malonate, a weaker chelator than citrate, Dy^{III} can be readily desorbed from both *Hans*-LanM and R100K with 10–100 mM chelator without significant Nd^{III} desorption, suggesting conditions for Nd^{III}/Dy^{III} separation (Fig. 4b).

Fig. 4: Leveraging *Hans*-LanM to separate Nd/Dy in a single-stage process.



a, *Hans*-LanM and the R100K variant exhibit greater differences in Nd versus Dy complex stability than *Mex*-LanM against desorption by citrate. Mean ± s.e.m. for three independent trials. **Significant difference between $[\text{citrate}]_{1/2}$ for La^{III} between *Hans*-LanM and *Hans*-LanM(R100K) (20 µM protein) shows the impact of dimerization of La^{III} complex stability ($P < 0.01$, analysis of variance with Bonferroni post-test). *Mex*-LanM Nd and Dy data from ref. [6]. **b**, Spectrofluorometric titration of *Hans*-LanM and R100K variant ($\lambda_{ex} = 280$ nm, $\lambda_{em} = 333$ nm) at pH 5.0, depicting the malonate-induced desorption of a 2:1 metal–protein complex. Mean ± s.e.m. for three independent trials, except those with R100K, which were single trials of each condition. **c**, Comparison of distribution factors (pH 5.0, about 0.33 mM each RE, La^{III}–Dy^{III}) for immobilized *Hans*-LanM, *Hans*-LanM(R100K) and *Mex*-LanM. Each point represents mean ± s.d. for three independent trials. **d**, Separation of a 95:5 mixture of Nd^{III}/Dy^{III} using

immobilized *Hans*-LanM(R100K) and a desorption scheme of three stepped concentrations of malonate followed by pH 1.5 HCl. One bed volume was 0.7 ml.

Source Data

Although a twofold modulation of RE versus RE selectivity by dimerization may seem small, such differences provide opportunity to decrease the number of separation stages, increasing efficiency of a separation process[3,12]. Therefore, *Hans*-LanM and the R100K variant were immobilized through a carboxy-terminal Cys residue on maleimide-functionalized agarose beads, as described previously[6], and tested for Nd^{III}/Dy^{III} separation. Immobilized *Hans*-LanM bound about one equivalent of RE, unlike in solution and compared with two equivalents for *Mex*-LanM[6] and *Hans*-LanM(R100K) (Supplementary Fig. 31). *Hans*-LanM and R100K exhibited similar separation ability in the La–Gd range—although R100K exhibits greater separation ability in the Gd–Dy range—as determined by the on-column distribution ratios (*D*) of a mixed RE solution at equilibrium (Fig. 4c, Extended Data Table 3 and Supplementary Tables 12–14). These Nd/Dy separation factors are nearly double (*Hans*-LanM) and triple (*Hans*-LanM(R100K)) that of *Mex*-LanM (Extended Data Table 3). Immobilized *Hans*-LanM was loaded to 90% of breakthrough capacity with a model electronic waste mixture of 5% dysprosium and 95% neodymium and, guided by Fig. 4b, eluted with a short, stepwise malonate gradient, followed by complete desorption using pH 1.5 HCl. In a single purification stage, Dy was upgraded from 5% to 83% purity and Nd was recovered at 99.8% purity (both >98% yield; Extended Data Fig. 6). This significantly outperformed the comparable *Mex*-LanM-based process, which achieved only 50% purity in a first separation stage and required a second stage to obtain >98% purity[6]. The immobilized R100K variant performed even better, achieving baseline separation of Dy^{III} and Nd^{III} to >98% purity and >99% yield in a single stage (Fig. 4d). The R100K variant's better performance was unexpected and may point to the unlikelihood of functional dimers on the column at this immobilization density (see the caption of Extended Data Fig. 6 for a discussion). Thus, despite substantially improved performance versus *Mex*-LanM enabled by characterization of *Hans*-LanM's mechanism of dimerization, fully exploiting the dimerization phenomenon

on-column may involve, for example, tethering of two monomers on a single polypeptide chain, which is under investigation.

Conclusion

Biochemical and structural characterization of *Hans*-LanM's mechanism of metal-sensitive dimerization provides a new, allosteric mechanism for LRE versus HRE selectivity in biology, extending concepts in dimer-dependent metal recognition recently emerging from synthetic lanthanide complexes[11] and engineered transition metal-binding proteins[40] and showing that these principles are hard-wired into nature. Our work also shows that dimerization strength, and thus metal selectivity, can be rationally modulated. *Hans*-LanM evolved LRE-selective dimerization at physiological protein concentrations closer to those in our biochemical assays (10–20 μM) rather than those on the column (about 3 mM). Therefore, leveraging dimerization in a separation process would be assisted by shifting dimerization sensitivity to the higher concentration regime, such as by tuning hydrophobic interactions at the dimerization interface. Furthermore, our studies establish that LanMs with as low as 33% identity are easily predicted yet have useful differences in metal selectivity; further mining of this diversity may reveal yet additional mechanisms for tuning RE separations. Finally, the solvent-excluded coordination spheres of *Hans*-LanM should outperform *Mex*-LanM in RE/actinide separation[23], luminescence-based sensing[21,26] and stabilization of hydrolysis-prone ions. Continued characterization of the coordination and supramolecular principles of biological f-element recognition will inspire design of ligands with higher RE versus RE selectivities and their implementation in new RE separation processes.

Methods

General considerations

See the Supplementary Methods for details.

Bioinformatics methods

Protein and genome sequence data

The sequence of LanM from *M. extorquens* AM1 was used as a query to conduct PSI-BLAST searches against the National Center for Biotechnology Information non-redundant protein sequence (nr) and metagenomic protein (env_nr) databases until convergence[41]. The resulting 3,047 protein sequences were then manually curated for those that are less than 200 residues long, have at least one pair of EF hands separated by less than 14 residues, and have 4 EF hands. Signal peptides of LanM sequences were predicted using SignalP (v6.0)[42], and then removed before further analysis of the sequences.

Construction of sequence similarity networks

The Enzyme Function Initiative-Enzyme Similarity Tool was used to calculate the similarities between all peptide sequence pairs with an *E*-value threshold of 1×10^{-5} (ref. [30]). The resulting sequence similarity network of 696 nodes and 241,853 edges was then constructed and explored using the organic layout through Cytoscape (v3.9.1)[43] and visualized in R (v4.1.0)[44]. The edge percentage identity threshold was gradually increased from 40% to 90% to yield distinct clusters.

Multiple sequence alignment and phylogenetic analysis

LanM sequences were aligned using MUSCLE (v5.1)[45] with default parameters. The model used for phylogeny construction was selected using ModelFinder in IQ-TREE (v2.2.0.3)[46,47] with --mset set to beast2. Bayesian phylogeny was generated on the basis of these results using BEAST (v2.6.7)[48]. The resulting phylogeny was evaluated using 10^7 generations and discarding a burn-in of 25%, and then visualized using ggtree (v3.2.1)[49].

Expression and purification of *Hans*-LanM and its R100K variant

The gene encoding *Hans*-LanM, codon optimized for expression in *E. coli* without its native 23-residue signal peptide (see Supplementary Table 15),

was obtained from Twist Bioscience and inserted into pET-29b(+) using the restriction sites NdeI/XhoI. *Hans*-LanM was overexpressed on a 2-l scale and purified using the established protocol for *Mex*-LanM[50], with one modification: after the final SEC step, the protein was concentrated to 5 ml and dialysed against 5 g Chelex 100 in 500 ml of 30 mM HEPES, 100 mM KCl, 5% glycerol, pH 8.4, to remove Ca^{II} and trace metal contaminants. This procedure resulted in approximately 15 ml of 550 µM protein, which was not concentrated further. The final yield was 45 mg of protein per litre of culture. Protein concentrations were calculated using an extinction coefficient of 11,000 M^{-1} cm^{-1}, based on the ExPASy ProtParam tool[51]. *Hans*-LanM(R100K) was purified using the same procedure, yielding 30 mg of protein per litre of culture.

Circular dichroism spectroscopy

Circular dichroism spectra of *Hans*-LanM were collected as described previously[32], at 15 µM (monomer concentration) in Chelex 100-treated buffer A (20 mM acetate, 100 mM KCl, pH 5.0), unless otherwise indicated. Buffered metal solutions were prepared as described previously[4,23,25,32]. Additional details are available in the Supplementary Information.

Preparation of protein samples for SEC–MALS and small-angle X-ray scattering (SAXS)

Samples of wild-type *Hans*-LanM were prepared by adding 3.0 equivalents of metal slowly (0.5 equivalent at a time followed by mixing) to 1.0 ml of concentrated stock of *Hans*-LanM (550 µM). At these protein concentrations, slight precipitation was observed for LRE samples (for example, La^{III}) whereas significant precipitation was observed for HRE samples (for example, Dy^{III}). Samples were centrifuged at 12,000*g* for 2 min to remove precipitate and then purified using gel filtration chromatography (HiLoad 10/300 Superdex 75 pg, 1-ml loop, 0.8 ml min^{-1}) in buffer B (30 mM MOPS, 100 mM KCl, 5% glycerol, pH 7.0). *Hans*-LanM-containing peaks (ranging from 12.0 to 15.0 ml elution volume) were collected to avoid the high-MW aggregate peaks, yielding 2.0 ml of metalated *Hans*-LanM ranging between 114 µM and 128 µM (1.37–1.53 mg ml^{-1}).

Samples of *Hans*-LanM(R100K) do not form high-MW species or precipitate on metal addition. To prepare samples of this protein, a 500 μM protein solution was diluted to 250 μM (3 mg ml^{-1}) in buffer B containing 0.75 mM of a specific $RECl_3$, yielding a final solution of 3 mg ml^{-1} protein, with a 1:3 metal ratio, which was analysed directly by SEC–MALS.

For calcium conditions, proteins were diluted to 250 μM (3 mg ml^{-1}), 5 mM $CaCl_2$ was added, and the samples were incubated at room temperature for 1 h. The buffer used for SEC–MALS was the same as above, except that it also contained 5 mM $CaCl_2$.

In-line SEC and MALS

SEC–MALS experiments were conducted using an Agilent 1260 Infinity II HPLC system equipped with an autosampler and fraction collector, and the Wyatt SEC hydrophilic column had 5-μm silica beads, a pore size of 100 Å and dimensions of 7.8 × 300 mm. Wyatt Technology DAWN MALS and Wyatt Optilab refractive index detectors were used for analysing the molar mass of peaks that eluted from the column. The SEC–MALS system was equilibrated for 5 h with buffer B. The system was calibrated with bovine serum albumin (monomer MW: 66 kDa) in the same buffer and normalization and alignment of the MALS and refractive index detectors were carried out. A volume of 15 μl of each sample was injected at a flow rate of 0.8 ml min^{-1} with a chromatogram run time of 25 min. Data were analysed using the ASTRA software (Wyatt). When small-angle X-ray scattering (SAXS) analysis was desired, a second run was carried out with 150 μl protein (about 4 mg ml^{-1}) injected, and 200-μl fractions of the main peak were collected. BioSAXS data were subsequently collected in triplicate.

Isothermal titration calorimetry

The dissociation constants for the dimers of apo, La^{III}-bound and Dy^{III}-bound *Hans*-LanM were determined by dilutive additions of a concentrated protein stock, followed using isothermal titration calorimetry on a TA Instruments Low-volume Auto Affinity isothermal titration calorimeter. The

syringe contained 300 µM protein (apo or 2 equivalents of Dy^{III} bound) or 150 µM or 540 µM (2 equivalents of La^{III} bound), and the cell contained 185 µl of a matched buffer (30 mM MOPS, 100 mM KCl, pH 7.0). Titrations were carried out at 30 °C. Titrations consisted of a first 0.2-µl injection followed by 17 × 2-µl injections, unless otherwise noted, with stirring at 125 r.p.m. and 180 s equilibration time between injections. The data were fitted using NanoAnalyze using the Dimer Dissociation model, yielding the dimer dissociation constant (K_{dimer}), enthalpy of dissociation (ΔH) and entropy of dissociation (ΔS). All parameters are shown in Extended Data Table 2.

K_{dimer} is defined as the dissociation constant for the equilibrium $D \rightleftharpoons 2M$, such that $K_{dimer} = [M]^2/[D]$, in which $[D]$ is the concentration of the dimer and $[M]$ is the concentration of the monomer, and the total protein concentration $[P]$ (as measured using the extinction coefficient for the monomer) is given by $[P] = [M] + 2[D]$. Therefore, $K_{dimer} = 2[M]^2/([P] - [M])$ or

$$2{[M]}^{2}+{K}_{{\rm{dimer}}}[M]-{K}_{{\rm{dimer}}}[P]=0$$
(1)

This equation can be used to estimate monomer and dimer concentrations during SEC–MALS experiments, using K_{dimer} values calculated from isothermal titration calorimetry experiments and $[P]$ from the SEC–MALS trace. This equation can also be used to estimate the maximum possible K_{dimer} for La^{III}-bound protein, given the SEC–MALS data.

SAXS

SAXS data were collected on RE-complexed *Hans*-LanM, at protein concentrations given in Supplementary Table 5 using equipment and under conditions described in the Supplementary Methods.

The forward scattering $I(0)$ and the radius of gyration (R_g) are listed in Supplementary Table 5 and were calculated using the Guinier approximation, which assumes that at very small angles ($q < 1.3/R_g$) the intensity is approximated as $I(q) = I(0)\exp[-1/3(qR_g)^2]$. In the La^{III}-, Nd^{III}-

and Dy^{III}-bound conditions, this agrees with the calculated size of 17.9 Å for the crystallographic dimer. The molecular mass was estimated using a comparison with SAXS data of a bovine serum albumin standard. The data files were analysed for Guinier R_g, maximum particle dimension (D_{max}), Guinier fits, Kratky plots and pair-distance distribution function using the ATSAS software[52]. GNOM, within ATSAS, was used to calculate the pair-distance distribution function $P(r)$, from which R_g and D_{max} were determined. Solvent envelopes were computed using DENSS[36]. The theoretical scattering profiles of the constructed models were calculated and fitted to experimental scattering data using CRYSOL[53]. OLIGOMER[54] was used to estimate the monomer and dimer fractions.

Preparation of protein samples for crystallography

To *Hans*-LanM (2 ml, 1.16 mM, buffer B), 3.0 equivalents of $LaCl_3$ or $DyCl_3$ were added slowly, 0.5 equivalents at a time with mixing, to minimize precipitation. Precipitate was removed by centrifugation at 12,000*g* for 2 min. Any soluble aggregates were removed and the protein was exchanged into buffer lacking glycerol (buffer C: 30 mM MOPS, 50 mM KCl, pH 7.0) by gel filtration chromatography (HiLoad 16/600 Superdex 75 pg, 1-ml loop, 0.75 ml min^{-1}). The peak in the 70–85 ml range was pooled, and the fractions were concentrated to about 500 µl with a final concentration of about 1.3 mM.

Mex-LanM was purified as described previously[50] and was exchanged into buffer C before crystallization. The protein was loaded with 3.5 equivalents of Nd^{III} ($NdCl_3$).

General crystallographic methods

Diffraction datasets were collected at the Life Sciences Collaborative Access Team ID-G beamline and processed with the HKL2000 package[55]. In all structures, phase information was obtained with phenix.autosol[56,57] through the single-wavelength anomalous diffraction method, in which lanthanide ions identified with HySS[58] were used as the anomalous scatterers. Initial

models were generated with phenix.autobuild[59] with subsequent rounds of manual modification and refinement in Coot[60] and phenix.refine[61]. In the final stages of model refinement, anisotropic displacement parameters and occupancies were refined for all lanthanide sites[62]. Model validation was carried out with the Molprobity server[63]. Figures were prepared using the PyMOL molecular graphics software package (Schrödinger, LLC).

La-bound *Hans*-LanM structure determination

Crystals were obtained by using the sitting drop vapour diffusion method, in which 1 µl of protein solution (15 mg ml^{-1}) was mixed with 1 µl 10 mM tri-sodium citrate, pH 7.0, and 27% (w/v) PEG 6000 in a 24-well plate from Hampton Research (catalogue number HR1-002) at room temperature. Thin plate-shaped crystals appeared in 3 days. Crystals suitable for data collection were mounted on rayon loops, soaked briefly in a cryoprotectant solution consisting of the well solution supplemented with 10% ethylene glycol, and flash-frozen in liquid N_2.

La^{III}-loaded *Hans*-LanM crystallized in the $P2_1$ space group ($\beta = 90.024°$) with four monomers in the asymmetric unit. The initial figure of merit and Bayesian correlation coefficient were 0.563 and 0.56, respectively[64]. The final model consists of residues 24–133 in each chain, 12 La^{III} ions (3 per chain in the first, second and third EF hands), 4 Na^{I} ions[65] (1 per chain in the fourth EF hand), 273 water molecules and 2 molecules of citrate. Of the residues modelled, 100% are in allowed or preferred regions as indicated by Ramachandran statistical analysis.

Dy-bound *Hans*-LanM structure determination

Crystals were obtained by using the sitting drop vapour diffusion method, in which 1 µl of protein solution (15 mg ml^{-1}) was mixed with 1 µl of 250 µM tri-sodium citrate, pH 7.0, and 27% (w/v) PEG 6000 in a 24-well plate from Hampton Research at room temperature. Thin plate-shaped crystals appeared within 1 month. Crystals suitable for data collection were mounted on rayon loops, soaked briefly in a cryoprotectant solution consisting of the

well solution supplemented with perfluoropolyether cryo oil from Hampton Research (catalogue number HR2-814) and flash-frozen in liquid N_2.

Dy^{III}-loaded *Hans*-LanM crystallized in the $P2_1$ space group ($\beta = 93.567°$) with four monomers in the asymmetric unit. The initial figure of merit and Bayesian correlation coefficient were 0.748 and 0.58, respectively[64]. The final model consists of residues 24–133 in each chain (except for chain D, for which residues 34–38 cannot be modelled), 14 Dy^{III} ions (4 in chains A and D, 3 in the second, third and fourth EF hands of chains B and C) and 656 water molecules. Of the residues modelled, 100% are in allowed or preferred regions as indicated by Ramachandran statistical analysis.

Collection of anomalous datasets is described in the Supplementary Methods.

Nd-bound *Mex*-LanM structure determination

Crystals were obtained by using the sitting drop vapor diffusion method, in which 1 µl of protein solution (35 mg ml^{-1}) was mixed with 1 µl of 0.1 M ammonium sulfate, 0.1 M Tris pH 7.5, and 20% (w/v) PEG 1500 in a 24-well plate from Hampton Research at room temperature. Thin plate-shaped crystals appeared within 6 months. Crystals suitable for data collection were mounted on rayon loops, soaked briefly in a cryoprotectant solution consisting of the well solution supplemented with perfluoropolyether cryo oil from Hampton Research and flash-frozen in liquid N_2.

Nd^{III}-loaded *Mex*-LanM crystallized in the $P2_12_12_1$ space group with one monomer in the asymmetric unit. The initial figure of merit and Bayesian correlation coefficient were 0.799 and 0.56, respectively[64]. The final model consists of residues 29–133, 4 Nd^{III} ions and 171 water molecules. Of the residues modelled, 100% are in allowed or preferred regions as indicated by Ramachandran statistical analysis.

Fluorescence spectroscopy

All fluorescence data were collected using a Fluorolog-QM fluorometer in configuration 75-21-C (Horiba Scientific) equipped with a double monochromator on the excitation arm and single monochromator on the emission arm. A 75-W xenon lamp was used as the light source for steady-state measurements and a pulsed xenon lamp was used for time-resolved measurements. Ten-millimetre quartz spectrofluorometry cuvettes (Starna Cells, 18F-Q-10-GL14-S) were used to collect data at 90° relative to the excitation path.

Fluorescence lifetime measurements were carried out using established methods[26,66]. In short, a solution of *Hans*-LanM with 2 equivalents of Eu^{III} added, totalling 4.5 ml, was prepared in 100% H_2O matrix (buffer: 25 mM HEPES, 75 mM KCl, pH 7.0). Half of this initial protein mixture (2.25 ml) was retained for future use and the remainder was exchanged to D_2O through lyophilization to remove H_2O and resuspension in 99.9% D_2O two times. The resulting protein solutions (in 100% H_2O and about 99% D_2O) were mixed in varying ratios to produce D_2O contents of 0%, 25%, 50% and 75%. The protein concentration was 20 µM. For each sample, the luminescence decay time constant (τ) was measured ($\lambda_{ex} = 394$ nm, $\lambda_{em} = 615$ nm) with 5,000 shots over a time span of 2,500 µs. τ was determined using the FelixFL Powerfit-10 software (Horiba Scientific) using a single exponential fit. $1/\tau$ was plotted against percentage composition of D_2O, and the slope of the resulting line (m) was determined. The q value was determined using the following equation from ref. [67]:

$$q=1.11[{\tau }_{{\rm{H}}2{\rm{O}}}^{-1}-{\tau }_{{\rm{D}}2{\rm{O}}}^{-1}-0.31+0.45{n}_{{\rm{OH}}}+0.99{n}_{{\rm{NH}}}+0.075{n}_{{\rm{O}}\mbox{--}{\rm{CNH}}}]$$
(2)

in which $\tau^{-1}{}_{H2O}$ and $\tau^{-1}{}_{D2O}$ are the inverses of the time constants in 100% H_2O and D_2O, respectively (the latter extrapolated using the equation of the fitted line), in ms^{-1}; and $n_{OH} = 0$, $n_{NH} = 0$, and $n_{O–CNH} = 1$ (resulting from the metal-coordinated Asn residues), on the basis of the *Hans*-LanM crystal structures. This equation simplifies to:

$$q=1.11[-m-0.31+0.075]$$
(3)

For fluorescence competition experiments, a solution of 20 μM *Hans*-LanM or the R100K variant was prepared in buffer A (pH 5.0) with two equivalents of metal (40 μM). Fluorescence emission spectra were collected with settings: λ_{ex} = 278 nm, λ_{em} 300–420 nm, integration time = 0.5 s, step size = 1 nm. Titrations were carried out through addition of at least 0.6 μl of titrant (from concentrated stock solutions of 10 mM–1 M citrate or malonate, pH 5.0). Spectra were corrected for dilution. Each experiment was carried out in triplicate.

Purification of Cys-containing variants

Hans-LanM(R100K)-Cys was expressed and purified as described for *Mex*-LanM-Cys (ref. [6]), with a final yield of 50 mg of protein per litre of culture. For *Hans*-LanM-Cys, the protein was purified by incorporating the same modifications from above, minus the dialysis step, to our previously described *Mex*-LanM-Cys purification, except that the SEC step was run using a reducing buffer (30 mM MOPS, 100 mM KCl, 5 mM TCEP, pH 7.0) with 5 mM EDTA, and frozen under liquid N_2 before immobilization.

Maleimide functionalization of agarose beads

The maleimide functionalization of amine-functionalized agarose beads was described previously[6]. See the Supplementary Information for complete details.

Immobilization of *Hans*-LanM and the R100K variant

Hans-LanM(R100K) immobilization was carried out using a thiol-maleimide conjugation reaction as described previously[6]. In the case of *Hans*-LanM, a final protein concentration of about 0.4 mM (8 ml) was combined with 1 ml of maleimide–microbeads and the conjugation reaction was carried out for 16 h at room temperature. Unconjugated *Hans*-LanM was removed by washing with coupling buffer, and the *Hans*-LanM microbeads were stored in coupling buffer for subsequent tests. To quantify

Hans-LanM immobilization yield, Pierce BCA Protein Assay (ThermoFisher Scientific) was used to determine the LanM concentration in the reaction solution before and after the conjugation reaction as previously described.

Batch experiment to determine separation factors

LanM-immobilized microbeads were washed with deionized water. Feed solution (5 ml, equimolar REs La–Dy, 3 mM total, pH 5.0) was added to 1 ml microbeads and incubated for 2 h. The liquid at equilibrium was collected and RE concentrations were determined by inductively coupled plasma mass spectrometry as $[M]_{ad}$. Then 4 ml of 0.1 M HCl was used to desorb REs from the microbeads and concentrations were measured by inductively coupled plasma mass spectrometry as $[M]_{de}$.

The RE distribution factor (D) between the LanM phase and the solution phase was calculated as:

$$D=\frac{{[M]}_{{\rm{LanM}}}}{{[M]}_{{\rm{Liquid}}}}$$
(4)

in which $[M]_{\rm{LanM}}$ and $[M]_{\rm{Liquid}}$ are the molar concentrations of each metal ion in the LanM phase and the solution phase at equilibrium, respectively. To account for the free liquid that absorbed on the agarose microbeads, the following correction was applied: $[M]_{\rm{Liquid}} = [M]_{ad}$; $[M]_{\rm{LanM}} = (4 \times [M]_{de} - [M]_{ad})/4$.

The separation factor is defined as:

$${\rm{SF}}=\frac{{D}_{{\rm{RE1}}}}{{D}_{{\rm{RE2}}}}$$
(5)

in which $D_{\rm{RE1}}$ and $D_{\rm{RE2}}$ are the distribution factors of RE1 and RE2, respectively.

Breakthrough column experiments

Columns were filled and run, and metal concentrations analysed, as described in our previous work[6]; details are available in the Supplementary Methods.

For the RE pair separation experiments, the metal ion purity and yield are defined as:

$${{\rm{Purity}}}_{{\rm{RE}}1}=\frac{{C}_{{\rm{RE}}1}}{{C}_{{\rm{RE}}1}+{C}_{{\rm{RE}}2}}$$
(6)

$${{\rm{Y}}{\rm{i}}{\rm{e}}{\rm{l}}{\rm{d}}}_{{\rm{R}}{\rm{E}}1}=\frac{{\rm{R}}{\rm{E}}1\,{\rm{r}}{\rm{e}}{\rm{c}}{\rm{o}}{\rm{v}}{\rm{e}}{\rm{r}}{\rm{e}}{\rm{d}}}{{\rm{T}}{\rm{o}}{\rm{t}}{\rm{a}}{\rm{l}}\,{\rm{R}}{\rm{E}}1\,{\rm{l}}{\rm{o}}{\rm{a}}{\rm{d}}{\rm{e}}{\rm{d}}}$$
(7)

in which C_{RE1} and C_{RE2} are the molar concentrations of RE1 and RE2, respectively.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data are available in the main text or the Supplementary Information. Coordinates have been deposited in the Protein Data Bank with accession codes 8DQ2 (La^{III}–*Hans*-LanM), 8FNR (Dy^{II}–*Hans*-LanM) and 8FNS (Nd^{III}–*Mex*-LanM). Source data are provided with this paper.

References

1. Cheisson, T. & Schelter, E. J. Rare earth elements: Mendeleev’s bane, modern marvels. *Science* **363**, 489–493 (2019).

2. Nash, K. L. & Jensen, M. P. Analytical-scale separations of the lanthanides: a review of techniques and fundamentals. *Sep. Sci. Technol.* **36**, 1257–1282 (2001).

3. Xie, F., Zhang, T. A., Dreisinger, D. & Doyle, F. A critical review on solvent extraction of rare earths from aqueous solutions. *Miner. Eng.* **56**, 10–28 (2014).

4. Cotruvo, J. A. Jr, Featherston, E. R., Mattocks, J. A., Ho, J. V. & Laremore, T. N. Lanmodulin: a highly selective lanthanide-binding protein from a lanthanide-utilizing bacterium. *J. Am. Chem. Soc.* **140**, 15056–15061 (2018).

5. Cook, E. C., Featherston, E. R., Showalter, S. A. & Cotruvo, J. A. Jr. Structural basis for rare earth element recognition by *Methylobacterium extorquens* lanmodulin. *Biochemistry* **58**, 120–125 (2019).

6. Dong, Z. et al. Bridging hydrometallurgy and biochemistry: a protein-based process for recovery and separation of rare earth elements. *ACS Cent. Sci.* **7**, 1798–1808 (2021).

7. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Cryst. A* **32**, 751–767 (1976).

8. Deblonde, G. J. P., Chagnes, A. & Cote, G. Recent advances in the chemistry of hydrometallurgical methods. *Sep. Purif. Rev.* https://doi.org/10.1080/15422119.2022.2088389 (2022).

9. Vahidi, E. & Zhao, F. Environmental life cycle assessment on the separation of rare earth oxides through solvent extraction. *J. Environ. Manage.* **203**, 255–263 (2017).

10. Jensen, M. P. et al. Solvent extraction separation of trivalent americium from curium and the lanthanides. *Solvent Extr. Ion Exch.* **33**, 329–345 (2015).

11. Bogart, J. A. et al. Accomplishing simple, solubility-based separations of rare earth elements with complexes bearing size-sensitive molecular

apertures. *Proc. Natl Acad. Sci. USA* **113**, 14887–14892 (2016).

12. Healy, M. R. et al. Efficient separation of light lanthanides(III) by using bis-lactam phenanthroline ligands. *Chem. Eur. J.* **25**, 6326–6331 (2019).

13. Thiele, N. A., Fiszbein, D. J., Woods, J. J. & Wilson, J. J. Tuning the separation of light lanthanides using a reverse-size selective aqueous complexant. *Inorg. Chem.* **59**, 16522–16530 (2020).

14. Slope, L. N., Daubney, O. J., Campbell, H., White, S. A. & Peacock, A. F. A. Location-dependent lanthanide selectivity engineered into structurally characterized designed coiled coils. *Angew. Chem. Int. Ed.* **60**, 24473–24477 (2021).

15. Fang, H. et al. Electro-kinetic separation of rare earth elements using a redox-active ligand. *Angew. Chem. Int. Ed.* **56**, 13450–13454 (2017).

16. Demir, S. et al. Extraction of lanthanide and actinide ions from aqueous mixtures using a carboxylic acid-functionalized porous aromatic framework. *ACS Cent. Sci.* **2**, 253–265 (2016).

17. Park, D. M. et al. Bioadsorption of rare earth elements through cell surface display of lanthanide binding tags. *Environ. Sci. Technol.* **50**, 2735–2742 (2016).

18. Deblonde, G. J.-P. et al. Selective and efficient biomacromolecular extraction of rare-earth elements using lanmodulin. *Inorg. Chem.* **59**, 11855–11867 (2020).

19. Lumpe, H. et al. The earlier the better: structural analysis and separation of lanthanides with pyrroloquinoline quinone. *Chem. Eur. J.* **26**, 10133–10139 (2020).

20. Ding, Y., Harvey, D. & Wang, N.-H. L. Two-zone ligand-assisted displacement chromatography for producing high-purity praseodymium, neodymium, and dysprosium with high yield and high productivity from crude mixtures derived from waste magnets. *Green Chem.* **22**, 3769–3783 (2020).

21. Deblonde, G. J. P. et al. Characterization of americium and curium complexes with the protein lanmodulin: a potential macromolecular mechanism for actinide mobility in the environment. *J. Am. Chem. Soc.* **143**, 15769–15783 (2021).

22. Deblonde, G. J.-P. et al. Capturing an elusive but critical element: natural protein enables actinium chemistry. *Sci. Adv.* **7**, eabk0273 (2021).

23. Mattocks, J. A., Cotruvo, J. A. Jr. & Deblonde, G. J. P. Engineering lanmodulin's selectivity for actinides over lanthanides by controlling solvent coordination and second-sphere interactions. *Chem. Sci.* **13**, 6054–6066 (2022).

24. Singer, H., Drobot, B., Zeymer, C., Steudtner, R. & Daumann, L. J. Americium preferred: lanmodulin, a natural lanthanide-binding protein favors an actinide over lanthanides. *Chem. Sci.* **12**, 15581–15587 (2021).

25. Mattocks, J. A., Ho, J. V. & Cotruvo, J. A. Jr. A selective, protein-based fluorescent sensor with picomolar affinity for rare earth elements. *J. Am. Chem. Soc.* **141**, 2857–2861 (2019).

26. Featherston, E. R., Issertell, E. J. & Cotruvo, J. A. Jr. Probing lanmodulin's lanthanide recognition via sensitized luminescence yields a platform for quantification of terbium in acid mine drainage. *J. Am. Chem. Soc.* **143**, 14287–14299 (2021).

27. Hussain, Z. et al. Repeated recovery of rare earth elements using a highly selective and thermo-responsive genetically encoded polypeptide. *Adv. Funct. Mater.* **32**, 2109158 (2022).

28. Agafonova, N. V., Kaparullina, E. N., Grouzdev, D. S. & Doronina, N. V. *Hansschlegelia quercus* sp. nov., a novel methylotrophic bacterium isolated from oak buds. *Int. J. Syst. Evol. Microbiol.* **70**, 4646–4652 (2020).

29. Renner, M., Danielson, M. A. & Falke, J. J. Kinetic control of Ca(II) signaling: tuning the ion dissociation rates of EF-hand Ca(II) binding sites. *Proc. Natl Acad. Sci. USA* **90**, 6493–6497 (1993).

30. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry* **58**, 4169–4182 (2019).

31. Lidstrom, M. E. in *The Prokaryotes* Vol. 2 (eds Dworkin, M. et al.) 618–634 (Springer, 2006).

32. Mattocks, J. A., Tirsch, J. L. & Cotruvo, J. A. Jr. Determination of affinities of lanthanide-binding proteins using chelator-buffered titrations. *Methods Enzymol.* **651**, 23–61 (2021).

33. Pol, A. et al. Rare earth metals are essential for methanotrophic life in volcanic mudpots. *Environ. Microbiol.* **16**, 255–264 (2014).

34. Wehrmann, M., Billard, P., Martin-Meriadec, A., Zegeye, A. & Klebensberger, J. Functional role of lanthanides in enzymatic activity and transcriptional regulation of pyrroloquinoline quinone-dependent alcohol dehydrogenases in *Pseudomonas putida* KT2440. *mBio* **8**, e00570-17 (2017).

35. Good, N. M. et al. Hyperaccumulation of gadolinium by *Methylorubrum extorquens* AM1 reveals impacts of lanthanides on cellular processes beyond methylotrophy. *Front. Microbiol.* https://doi.org/10.3389/fmicb.2022.820327 (2022).

36. Grant, T. D. Ab initio electron density determination directly from solution scattering data. *Nat. Methods* **15**, 191–193 (2018).

37. Deng, Y. W., Ro, S. Y. & Rosenzweig, A. C. Structure and function of the lanthanide-dependent methanol dehydrogenase XoxF from the methanotroph *Methylomicrobium buryatense* 5GB1C. *J. Biol. Inorg. Chem.* **23**, 1037–1047 (2018).

38. Habenschuss, A. & Spedding, F. H. The coordination (hydration) of rare earth ions in aqueous chloride solutions from x-ray diffraction. II. $LaCl_3$, $PrCl_3$, and $NdCl_3$. *J. Chem. Phys.* **70**, 3758–3763 (1979).

39. Hu, A., MacMillan, S. N. & Wilson, J. J. Macrocyclic ligands with an unprecedented size-selectivity pattern for the lanthanide ions. *J. Am. Chem. Soc.* **142**, 13500–13506 (2020).

40. Choi, T. S. & Tezcan, F. A. Overcoming universal restrictions on metal selectivity by protein design. *Nature* **603**, 522–527 (2022).

41. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402 (1997).

42. Teufel, F. et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **40**, 1023–1025 (2022).

43. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).

44. R Core Team. R: *A Language and Environment for Statistical Computing*. http://www.R-project.org/ (R Foundation for Statistical Computing, Vienna 2018).

45. Edgar, R. C. High-accuracy alignment ensembles enable unbiased assessments of sequence homology and phylogeny. *Nat. Commun.* https://doi.org/10.1038/s41467-022-34630-w (2022).

46. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).

47. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

48. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).

49. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

50. Featherston, E. R., Mattocks, J. A., Tirsch, J. L. & Cotruvo, J. A. Jr. Heterologous expression, purification, and characterization of proteins in the lanthanome. *Methods Enzymol.* **650**, 119–157 (2021).

51. Gasteiger, E. et al. in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) 571–607 (Humana, 2005).

52. Manalastas-Cantos, K. et al. ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Cryst.* **54**, 343–355 (2021).

53. Franke, D. et al. ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Cryst.* **50**, 1212–1225 (2017).

54. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).

55. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

56. McCoy, A. J., Storoni, L. C. & Read, R. J. Simple algorithm for a maximum-likelihood SAD function. *Acta Crystallogr. D* **60**, 1220–1228 (2004).

57. Zwart, P. H. et al. Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435 (2008).

58. Grosse-Kunstleve, R. W. & Adams, P. D. Substructure search procedures for macromolecular structures. *Acta Crystallogr. D* **59**, 1966–1973 (2003).

59. Terwilliger, T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D* **64**, 61–69 (2008).

60. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

61. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).

62. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).

63. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).

64. Doutch, J., Hough, M. A., Hasnain, S. S. & Strange, R. W. Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. *J. Synchrotron Rad.* **19**, 19–29 (2012).

65. Zheng, H. et al. CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr. D* **73**, 223–233 (2017).

66. Chaudhuri, D., Horrocks, W. D., Amburgey, J. C. & Weber, D. J. Characterization of lanthanide ion binding to the EF-hand protein S100β by luminescence spectroscopy. *Biochemistry* **36**, 9674–9680 (1997).

67. Supkowski, R. M. & Horrocks, W. D. On the determination of the number of water molecules, *q*, coordinated to europium(III) ions in solution from luminescence decay lifetimes. *Inorg. Chim. Acta* **340**, 44–48 (2002).

68. Seitz, M., Oliver, A. G. & Raymond, K. N. The lanthanide contraction revisited. *J. Am. Chem. Soc.* **129**, 11153–11160 (2007).

69. Cramer, R. E., Rimsza, J. M. & Boyle, T. J. The lanthanide contraction is a variable. *Inorg. Chem.* **61**, 6120–6127 (2022).

70. Drake, S. K., Lee, K. L. & Falke, J. J. Tuning the equilibrium ion affinity and selectivity of the EF-hand calcium binding motif: substitutions at the gateway position. *Biochemistry* **35**, 6697–6705 (1996).

Acknowledgements

This work was financially supported primarily by Department of Energy (DOE) grant DE-SC0021007 (to J.A.C.), as well as by NSF grant CHE-1945015 (to J.A.C.), supporting the *Hans*-LanM and *Mex*-LanM work, respectively. A.K.B. acknowledges National Institutes of Health grant GM119707 and C.-Y.L. acknowledges a fellowship from the Jane Coffin Childs Memorial Fund for Medical Research. D.M.P., C.S.K.-Y. and Z.D. acknowledge financial support from the Critical Materials Institute, an Energy Innovation Hub funded by the DOE, Office of Energy Efficiency and Renewable Energy, Advanced Materials and Manufacturing Technologies Office. Part of this work was carried out under the auspices of the DOE by Lawrence Livermore National Laboratory under contract DEAC52-07NA27344 (LLNL-JRNL-840467). This research used resources of the Advanced Photon Source, a DOE Office of Science User Facility operated by Argonne National Laboratory under contract number DE-AC02-06CH11357. Use of the Life Sciences Collaborative Access Team Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817). N.H.Y. acknowledges National Institutes of Health SIG awards S10-OD0215145 for isothermal titration calorimetry instrumentation, S10-OD028589 for the SAXS instrumentation and S10-OD030490 for the SEC–MALS–dynamic light scattering system. We thank J. Fecko for experimental assistance and Y. Jiao and G. Deblonde for discussions.

Author information

Authors and Affiliations

1. Department of Chemistry, The Pennsylvania State University, University Park, PA, USA

   Joseph A. Mattocks, Jonathan J. Jung, Chi-Yun Lin, Emily R. Featherston, Timothy A. Hamilton, Amie K. Boal & Joseph A. Cotruvo Jr

2. Critical Materials Institute, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, Livermore, CA, USA

   Ziye Dong, Christina S. Kang-Yun & Dan M. Park

3. The Huck Institutes of the Life Sciences, The Pennsylvania State University, University Park, PA, USA

   Neela H. Yennawar

4. Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA, USA

   Amie K. Boal

Contributions

J.A.C. identified *Hans*-LanM and conceived and directed the study. J.A.M. purified *Hans*-LanM and carried out most biochemical analyses, with assistance from T.A.H. J.J.J. crystallized the proteins, and J.J.J. and C.-Y.L. solved the structures with input from A.K.B. N.H.Y. carried out SAXS studies and supervised some biophysical data collection. E.R.F. purified *Mex*-LanM. C.S.K.-Y. carried out bioinformatics analyses. Z.D. carried out metal separation experiments, with input from D.M.P. J.A.M., N.H.Y., A.K.B. and J.A.C. wrote the paper with inputs from all authors. All authors edited and approved the final manuscript.

Corresponding authors

Correspondence to Dan M. Park, Amie K. Boal or Joseph A. Cotruvo Jr.

Ethics declarations

Competing interests

J.A.M., J.J.J., C.-Y.L., Z.D., E.R.F., C.S.K.-Y., D.M.P., A.K.B. and J.A.C. are inventors on a patent application filed by The Pennsylvania State University based on this work.

Peer review

Peer review information

Nature thanks Scott Banta and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Expanded view of the inset from Fig. 1a (*Hans* cluster) including 20 sequences and 190 edges.

The *Hans* cluster includes LanMs from bacteria from genera *Hansschlegelia*, *Ancylobacter*, *Methylopila*, *Oharaeibacter*, *Starkeya*, and *Xanthobacter*. Although these genera are restricted to this cluster, members at the family level are found dispersed throughout the network, including one *Xanthobacteraceae* and 42 *Methylocystaceae*.

Extended Data Fig. 2 Circular dichroism titrations of *Hans*-LanM with chelator-buffered solutions of (a) Ca^{II}, (b) Nd^{III}, and (c) Dy^{III}.

Both Dy^{III} (up to 0.3 μM) and Ca^{II} (up to 5.5 mM) induce a similar, incomplete conformational change in the protein, relative to the conformational change induced by Nd^{III} and La^{III}. The data in the right panel of **a** is a representative titration from the 3 datasets used to generate the plot in the left panel. The data in **b** and **c** are representative titrations from the 3 datasets used to generate the plot in Fig. 1d. Conditions: 15 μM protein, 20 mM acetate, 100 mM KCl, 10 mM EDTA (for Ca and Nd titrations) or EGTA (for Dy titration), 0—10 mM metal ion. Each data point in (**a**, left panel) is the mean ± s.d. for three independent measurements.

Source Data

Extended Data Fig. 3 X-ray crystal structure of La^{III}-bound *Hans*-LanM, solved at 1.8 Å resolution.

a, Overall structure of the asymmetric unit, which consists of two *Hans*-LanM dimers and two citrate molecules from the crystallization solution. The structure of each monomer of the dimer is consistent with the NMR solution structure of Y^{III}-bound *Mex*-LanM with EF-hands 2 and 3 paired and EF-hands 1 and 4 paired[5]. **b–e**, Details of metal coordination in the four EF-hands of La^{III}-*Hans*-LanM. The coordination spheres of the La^{III} ions in EF-hands 1, 2, and 3 are constituted by the side chain Oδ of N_1 (monodentate), the carboxylate side chains from D_3, D_5, E_9, and E_{12} (all bidentate), and a backbone carbonyl from S_7 (EF1) or T_7 (EF2 and EF3), for a total coordination number of 10. All La^{III}-ligand distances are 2.5-2.7 Å. The crystal radius for 10-coordinate La^{III} is given as 1.41 Å by Shannon;[7] given 1.26 Å as radius of 6-coord O^{2-}, 2.57 Å is estimated for the La^{III}-O distance, consistent with our results. The metal ion in EF4 was modeled as Na^{I} because of the shorter metal-ligand distances, lower coordination number, and the presence of sodium in the crystallization solution. Ca^{II}

cannot be completely ruled out as it was present earlier in the protein purification; however, the protein was treated with Chelex at the end of the purification, and the crystallographic data were consistent with the Na^{I} assignment as determined by the CheckMyMetal server[65]. This ion is coordinated in distorted pentagonal bipyramidal geometry by monodentate D_1, N_3, and D_5 sidechains, the bidentate E_{12} sidechain, the backbone carbonyl of Lys113, and a single solvent molecule for a total coordination number of 7. The Na^{I}-protein ligand distances are 2.3-2.5 Å, with a solvent molecule at 2.7 Å. In the case of *Mex*-LanM, biochemical data[4,26] and NMR spectroscopy[5] have also supported EF4 as a poor lanthanide-binding site, and it was modeled without a metal ion in the NMR solution state structure[5].

Extended Data Fig. 4 X-ray crystal structure of Dy^{III}-bound *Hans*-LanM, solved at 1.4 Å resolution.

a, One of the dimers in the asymmetric unit, comprising chains A and B. Note that EF4 is unexpectedly occupied with Dy^{III} while EF1 is occupied only in chain A. **b**, Overall structure of the asymmetric unit, which consists of two *Hans*-LanM dimers. Unlike the La^{III}-*Hans*-LanM structure, the two dimers – and the monomers within each dimer – display significant differences in Dy^{III}-*Hans*-LanM. EF2-4 are occupied by Dy^{III} in all chains, whereas EF1 is only occupied and ordered in chain A; in chains B and C, no metal ion is bound in the EF-hand, and in chain D, a Dy^{III} ion is bound but the first five residues of EF1 (Asn34 – Asp38) could not be modeled. Our decision to model Dy^{III} into all four EF-hands is supported by anomalous diffraction datasets (Supplementary Tables 9–10, Supplementary Figs. 25–26). The biochemical data suggest that, in solution, at least one Dy^{III} binding site is weak (see Fig. 1d and Supplementary Fig. 2), and it is likely based on studies of *Mex*-LanM that EF2/3 are the tighter binding sites. This proposal is supported by the Dy anomalous data (Supplementary Table 10), and the occupancy of weak metal-binding sites likely results from the high protein concentration used for crystallography. **c**, Details of metal coordination in the EF-hands of Dy^{III}-*Hans*-LanM. In the top row, the three different EF1 structures in the asymmetric unit are shown. Only in chain A is the EF1 metal site nearly identical to the sites in EF2 and EF3 (contrary to La^{III}-

Hans-LanM, where EF1-3 sites are very similar, Extended Data Fig. 3). In EF1 (chain A), EF2, and EF3, the coordinating ligands are the same as with La^{III}-*Hans*-LanM, except that the E_9 residues (Glu42, Glu66, and Glu91) have shifted to monodentate coordination, resulting in 9-coordination. The lower coordination number with Dy^{III} is consistent with the lanthanide contraction[68,69] and is observed with other ligands (as one recent example, ref. [39]). The Dy^{III}-ligand distances are mostly 2.3-2.5 Å, ~0.2 Å shorter than for La^{III}-*Hans*-LanM. Consistent with this observation, the crystal radius for 9-coordinate Dy^{III} is given as 1.22 Å by Shannon[7], 0.19 Å shorter than for 10-coordinate La^{III} (Extended Data Fig. 3). The carboxylate shift of the 9th position Glu residue is noteworthy as this position is important for gating affinity and selectivity in other EF-hand proteins[70]. In EF4, Dy^{III} is 7-coordinate with pentagonal bipyramidal geometry, similar to the sodium site in La^{III}-*Hans*, but with slightly shorter metal-ligand distances (2.2-2.5 Å); again, these distances are consistent with the expectation for 7-coordinate Dy^{III} (ref. [7]).

Extended Data Fig. 5 Spectrofluorometric titrations of RE-LanM (*Hans*-LanM, *Hans*-LanM(R100K), and *Mex*-LanM) complexes with citrate as a competitor, monitored by intrinsic protein fluorescence.

Emission values are normalized to 1.0 for the fluorescence of the apoprotein. Note that the fluorescence intensity of *Hans*-LanM's Trp residues decreases going from the RE-bound to apo state (Supplementary Fig. 30), whereas the intensity of *Mex*'s Tyr residue increases going from the RE-bound to apo state[4,6]. Initial conditions: 20 μM protein, 40 μM RE, 20 mM acetate, 100 mM KCl, pH 5.0, for all experiments, into which increasing concentrations of citrate were titrated. The citrate concentrations at which 50% of each metal is desorbed under these conditions ($[citrate]_{1/2}$) are summarized in Supplementary Table 11 and plotted in Fig. 4a. **a**, *Hans*-LanM. **b**, *Hans*-LanM(R100K). The compressed difference between the $[citrate]_{1/2}$ values for La and Nd in *Hans*-LanM(R100K) vs. wild-type *Hans*-LanM illustrates the role of dimerization in enhancing affinity differences for the LREs, especially La^{III}. **c**, *Mex*-LanM. Nd and Dy data were reported in Dong et al[6].

d, Comparison of the ratios of $[\text{citrate}]_{1/2}$ value for Nd to that for Dy, for each protein, illustrating the greater Nd/Dy selectivity of *Hans*-LanM relative to *Mex*-LanM. The ratios for wild-type *Hans*-LanM and the R100K variant are not significantly different ($p > 0.05$) by two-tailed t-test, suggesting that the hydrogen-bonding network involving Arg100 contributes relatively little to Nd^{III}/HRE selectivity, though it does impact La^{III} selectivity significantly (Fig. 4a). All data are shown as mean ± s.d. (a—c) or s.e.m. (d) for data from 3 independent experiments.

Source Data

Extended Data Fig. 6 Separation of a 95:5 mixture of Nd:Dy using immobilized *Hans*-LanM.

The desorption scheme consisted of three stepped concentrations of malonate (30, 50, 90 mM; see right axis) followed by pH 1.5 HCl. The results revealed that slightly lower purity Dy was generated using *Hans*-LanM compared to the R100K variant (83.6% vs 98% Dy purity at similar yield, respectively; compare to Fig. 4d). While similar selectivity profiles were observed for the immobilized proteins for La through Gd in equilibrium binding experiments with La-Dy, the selectivity pattern diverged at Tb (Fig. 4c). The selectivity difference between *Hans*-LanM and the R100K variant was confirmed by using a Nd/Dy binary system, as the uncertainties in the distribution factor determination for Dy in the 9-element RE group precluded the ability to distinguish small differences in the Dy/Nd separation factor between proteins (Supplementary Tables S13–S14). In this binary Nd/Dy experiment (Extended Data Table 3), we determined a separation factor of 8.12 ± 0.40 for *Hans*-LanM and 12.7 ± 1.3 for the R100K variant, which is consistent with the improved Dy separation efficacy of R100K. While consistent with the values derived from the 9-element experiment, the results differ slightly from the equilibrium binding results with the free *Hans*-LanM and *Hans*-LanM(R100K) proteins, which revealed similarly high selectivity for Nd over Dy (Fig. 4a,b), likely reflecting weaker LRE-induced dimerization in the R100K variant at the low protein concentration (20 μM) of the solution experiments with free protein. The La/Nd selectivity on-column is also distinct from that observed with the apparent K_d values of the free proteins (wild-type and R100K) in solution,

although the experiments with free proteins utilized single element solutions and effects from mixed metal binding may impact the on-column data. The R100K variant is also better behaved on the column, as evidenced by the 2:1 RE:protein stoichiometry. One possible explanation for these results could be that immobilization interferes with dimerization; however, Fig. 2b shows the N- and C-termini of the *Hans*-LanM dimer, indicating that the C-termini are ~20 Å from the nearest part of the dimer interface, suggesting that immobilization per se would not be expected to disrupt this interface. It must be considered, however, that a functional dimer would require two C-termini to be immobilized in close proximity, which is unlikely at the immobilization densities of our columns. Therefore, on balance, we suspect that the dimerization equilibrium is only applicable in a minority of protein units immobilized on the column. We posit that more fully exploiting the dimerization equilibrium in the column format would yield even more robust separations. The surest way to obtain homogeneous populations of dimers on-column would likely be to link two monomers together (e.g., with a polypeptide chain), tuning dimerization affinity through mutagenesis of the residues contributing to inter-monomer interactions, and immobilizing this dimer through a single attachment point. Dimerization could also be exploited in other separation formats. These directions are the subject of current efforts.

Source Data

Extended Data Table 1 Characterization of *Hans*- and *Mex*-LanM metal complexes using SEC-MALS
Extended Data Table 2 Thermodynamic parameters for apo- and Dy^{III}-bound *Hans*-LanM, obtained by fitting ITC thermograms to the dimer dissociation model
Extended Data Table 3 Distribution factors (*D*) and separation factors (SF) for equilibration of a binary Nd/Dy solution using a *Hans*-LanM column or *Hans*-LanM(R100K) column

Supplementary information

Supplementary Information

This file contains Supplementary Methods, Supplementary Figs 1-31, Supplementary Tables 1-16 and Supplementary References.

Reporting Summary

Source Data for Supplementary Fig. 3

Source Data for Supplementary Fig. 10

Source Data for Supplementary Fig. 19

Source Data for Supplementary Fig. 21

Source Data for Supplementary Fig. 30

Source Data for Supplementary Table 12

Source Data for Supplementary Table 13

Source Data for Supplementary Table 14

Source data

Source Data Fig. 1

Source Data Fig. 4

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Table 3

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Mattocks, J.A., Jung, J.J., Lin, CY. *et al.* Enhanced rare-earth separation with a metal-sensitive lanmodulin dimer. *Nature* **618**, 87–93 (2023). https://doi.org/10.1038/s41586-023-05945-5

- Received: 29 December 2022
- Accepted: 13 March 2023
- Published: 31 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05945-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

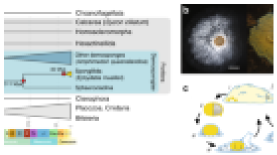
Provided by the Springer Nature SharedIt content-sharing initiative

How a protein differentiates between rare-earth elements

- Scott Banta

Nature News & Views 31 May 2023

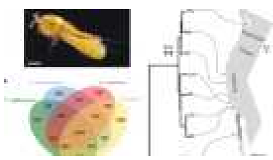
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05945-5



Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge Ephydatia muelleri

27 July 2020

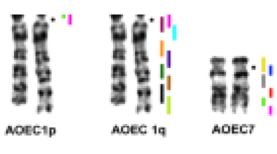
Nathan J. Kenny, Warren R. Francis, … Sally P. Leys



Genomic signatures of mitonuclear coevolution across populations of Tigriopus californicus

09 July 2018

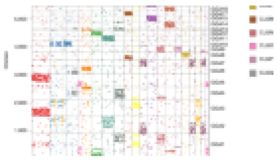
Felipe S. Barreto, Eric T. Watson, … Ronald S. Burton



Multiple intrasyntenic rearrangements and rapid speciation in voles

08 October 2018

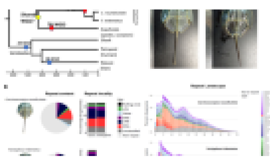
Svetlana A. Romanenko, Natalya A. Serdyukova, … Alexander S. Graphodatsky



[Deeply conserved synteny resolves early events in vertebrate evolution]

20 April 2020

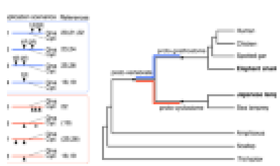
Oleg Simakov, Ferdinand Marlétaz, … Daniel S. Rokhsar



[Horseshoe crab genomes reveal the evolution of genes and microRNAs after three rounds of whole genome duplication]

19 January 2021

Wenyan Nong, Zhe Qu, … Jerome H. L. Hui



[Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution]

23 July 2021

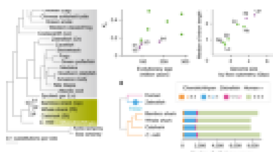
Yoichiro Nakatani, Prashant Shingate, … Byrappa Venkatesh



[The complete mitochondrial genome of the eusocial sponge-dwelling snapping shrimp Synalpheus microneptunus]

08 May 2020

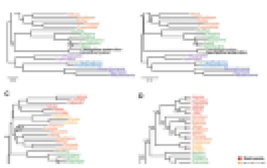
Solomon T. C. Chak, Phillip Barden & J. Antonio Baeza



[Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates]

08 October 2018

Yuichiro Hara, Kazuaki Yamaguchi, … Shigehiro Kuraku



[New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships]

09 January 2020

Kevin M. Kocot, Albert J. Poustka, … Michael Schrödl

- Article
- Open Access
- Published: 17 May 2023

Ancient gene linkages support ctenophores as sister to other animals

- Darrin T. Schultz ,
- Steven H. D. Haddock,
- Jessen V. Bredeson,
- Richard E. Green,
- Oleg Simakov &
- …
- Daniel S. Rokhsar

Show authors

Nature **volume 618**, pages 110–117 (2023)

- 37k Accesses
- 1654 Altmetric
- Metrics details

Abstract

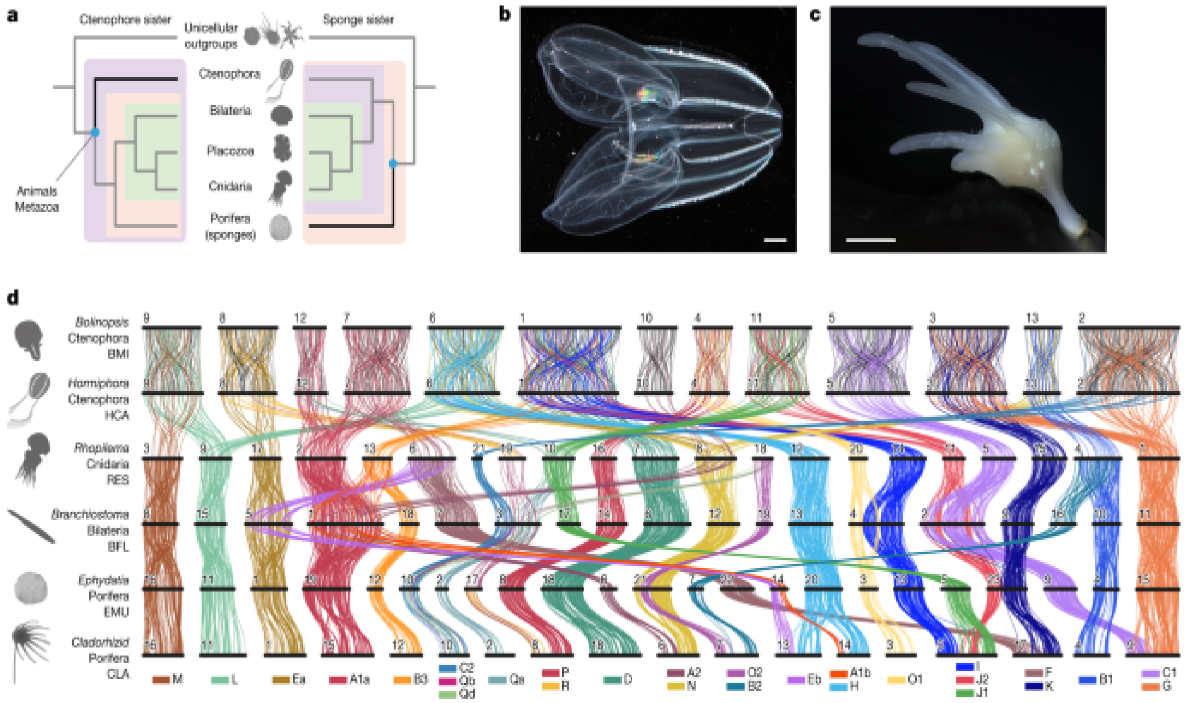
A central question in evolutionary biology is whether sponges or ctenophores (comb jellies) are the sister group to all other animals. These alternative phylogenetic hypotheses imply different scenarios for the evolution of complex neural systems and other animal-specific traits[1,2,3,4,5,6]. Conventional phylogenetic approaches based on morphological characters and increasingly extensive gene sequence collections have not been able to definitively answer this question[7,8,9,10,11]. Here we develop chromosome-scale gene linkage, also known as synteny, as a phylogenetic character for resolving this question[12]. We report new chromosome-scale genomes for a ctenophore and two marine sponges, and for three unicellular relatives of

animals (a choanoflagellate, a filasterean amoeba and an ichthyosporean) that serve as outgroups for phylogenetic analysis. We find ancient syntenies that are conserved between animals and their close unicellular relatives. Ctenophores and unicellular eukaryotes share ancestral metazoan patterns, whereas sponges, bilaterians, and cnidarians share derived chromosomal rearrangements. Conserved syntenic characters unite sponges with bilaterians, cnidarians, and placozoans in a monophyletic clade to the exclusion of ctenophores, placing ctenophores as the sister group to all other animals. The patterns of synteny shared by sponges, bilaterians, and cnidarians are the result of rare and irreversible chromosome fusion-and-mixing events that provide robust and unambiguous phylogenetic support for the ctenophore-sister hypothesis. These findings provide a new framework for resolving deep, recalcitrant phylogenetic problems and have implications for our understanding of animal evolution.

Main

Five major lineages arose early in animal evolution and survive to the present day: sponges (poriferans), ctenophores (comb jellies), placozoans (microscopic flat animals), cnidarians (such as anemones, jellyfishes and hydra) and bilaterians (such as chordates, molluscs, arthropods and diverse worms)[1,8,10,13,14]. Although morphological and phylogenomic studies consistently unite bilaterians, cnidarians, and placozoans into a monophyletic clade (Parahoxozoa) that excludes sponges and ctenophores[8,10,14] the relationship between sponges, ctenophores and Parahoxozoa remains controversial. There are two competing scenarios—the sponge-sister hypothesis[7,8] and the ctenophore-sister hypothesis[9,10]—reflecting which lineage diverged first among animals (Fig. 1a).

Fig. 1: Conserved synteny and the phylogenetic position of ctenophores and sponges.



a, Two alternative metazoan phylogenetic hypotheses, with either ctenophores (left) or sponges (right) as sister to all other animals. **b**,**c**, Specimens of species of which the genomes are reported here. Scale bars, 1 cm. **b**, The lobate ctenophore *B. microptera* from the Monterey Bay, California. **c**, Undescribed cladorhizid demosponge collected offshore of Big Sur, California at a depth of 3,975 m. **d**, Ribbon diagram showing conserved syntenies among animals ($\alpha \leq 0.05$, permutation test one-sided false-discovery rate), including (from top to bottom) two ctenophores (*B. microptera* (BMI) and *H. californensis*); the jellyfish *R. esculentum*; the bilaterian amphioxus *B. floridae*; and two demosponges (*E. muelleri* and the cladorhizid demosponge). Each horizontal black bar represents a chromosome. The vertical lines between species represent orthologous genes, coloured according to the BCnS synteny groups[12]. Only groups of genes that have significantly conserved chromosome-scale linkage (synteny) between metazoan species are shown. There is extensive 1:1 conserved chromosomal synteny between the two ctenophores, consistent with the conserved ctenophore karyotype. Orthologous gene pairs in the two ctenophores that do not participate in conserved syntenies with BCnS are shown in grey. Photography credits: Shannon Johnson © 2019 MBARI (**b**), © 2021 MBARI (**c**).

As sponges and ctenophores are such disparate animals[13], the nature of the first diverging animal lineage has implications for the evolution of fundamental animal characteristics. Adult sponges are generally sessile filter-feeding organisms with body plans organized into reticulated water-filtration channels, structures built out of silica or calcium carbonate, and specialized cell types and tissues used for feeding, reproduction and self-defence, but they lack neuronal and muscle cells[15]. By contrast, ctenophores are gelatinous marine predators that move using eight longitudinal ‘comb rows’ of ciliary bundles[16,17]; they are superficially similar but unrelated to cnidarian medusae[13,18] and possess multiple nerve nets[19]. Thus, whereas the sponge-sister scenario suggests a single origin of neurons on the ctenophore–parahoxozoan stem, the ctenophore-sister scenario implies either that either ancestral metazoan neurons were lost in the sponge lineage, or that there was convergent evolution of neurons in the ctenophore and parahoxozoan lineages[3,6]. Similar considerations apply to other metazoan cell types[18], gene regulatory networks, animal development[13,18] and other uniquely metazoan features.

Despite its importance for understanding animal evolution, the relative branching order of sponges, ctenophores and other animals has proven to be difficult to resolve[2]. The fossil record is largely silent on this issue as verified Precambrian sponge fossils are extremely rare[20] and putative fossils of the soft-bodied ctenophores are difficult to interpret[21]. Morphological characters of living groups (for example, choanocytes of sponges) are not sufficient to resolve the question because true homology is difficult to assign, and such characters are easily lost or can arise convergently[13,22]. The ctenophore-sister hypothesis is supported by a pair of gene duplications shared by sponges, bilaterians, placozoans and cnidarians but not ctenophores[23]. Although sophisticated methods for sequence-based phylogenomics have been developed and applied to increasingly large molecular datasets, there is still considerable debate about the relative position of sponges and ctenophores as results are sensitive to how sequence evolution is modelled[11], which taxa or sites are included[24,25], and the effects of long-branch artifacts and nucleotide compositional variation[26]. New approaches are needed.

We reasoned that patterns of synteny, classically defined as chromosomal gene linkage without regard to gene order[27], could provide a powerful tool for resolving the ctenophore-sister versus sponge-sister debate. Chromosomal patterns of gene linkage evolve slowly in many lineages[12,28,29,30], probably because it is improbable for interchromosomal translocations to be fixed in populations with large effective population sizes[28,31,32]. Notably, some changes in synteny are effectively irreversible. For example, when two distinct ancestral synteny groups are combined onto a single chromosome by translocation, and subsequent intrachromosomal rearrangements mix these two groups of genes, it is very unlikely that the ancestral separated pattern will be restored by further rearrangement and fission, in the same sense that spontaneous reduction in entropy is improbable[12]. Such rare and irreversible changes are particularly useful for resolving challenging phylogenetic questions as they give rise to shared derived features that unambiguously unite all descendant lineages[33,34,35]. Deeply conserved syntenies observed between animals and their closest unicellular relatives[12] suggest that outgroup comparisons could be used to infer ancestral metazoan states and polarize changes within animals to address the sponge-sister versus ctenophore-sister debate. Yet, chromosome-scale genome sequences of the unicellular or colonial eukaryotic outgroups closest to animals (choanoflagellates, filastereans and ichthyosporeans) have not been reported.

Here we show that conserved syntenies between animals and their closest unicellular relatives support ctenophores as the sister group to all other animals. Specifically, we find seven sets of genes for which (1) ctenophores share ancestral metazoan gene linkages with one or more unicellular eukaryotes; and (2) bilaterians, cnidarians, placozoans and sponges are united (to the exclusion of ctenophores) by shared derived patterns of synteny that arose by ancient interchromosomal translocations. In four of these cases, irreversible mixing after chromosome fusion evidently occurred on the bilaterian–cnidarian–sponge (BCnS) stem lineage, providing unambiguous support for the ctenophore-sister scenario. The alternative sponge-sister hypothesis is not supported by any synteny-based characters, and would require reversal of four sets of fusion-with-mixing events and/or extensive convergent fusion in both sponges and on the bilaterian/cnidarian stem to account for the observed patterns of synteny. To enable these

analyses, we generated chromosome-scale genome sequences for three animal species (two sponges and a ctenophore), and three non-animal species (a filasterean, ichthyosporean and choanoflagellate) to serve as outgroups. Our analyses further reveal ancient syntenies conserved between animals and their closest unicellular relatives (animal plesiomorphies) as well as metazoan syntenies shared by all animals but not present in unicellular organisms (animal synapomorphies). These findings establish a phylogenetic framework for understanding the early evolution of metazoan genomes and characters.

To examine conserved syntenies across animals, we traced the chromosomal distribution of orthologous genes among diverse metazoan lineages using previously and newly sequenced genomes (Fig. 1bc, Methods, Supplementary Information 1–3 and Supplementary Data 1). Figure 1d highlights conserved metazoan synteny groups, that is, groups of genes of which orthologues are linked on the same chromosome across multiple lineages, regardless of gene order. Syntenic groups shown in Fig. 1d are statistically significant (Methods). In Fig. 1d, lines connecting orthologous genes are coloured according to the previously identified BCnS ancestral linkage groups (ALGs)[12,28,29,30]. For example, the group on the far left represents the BCnS ALG_M (comprising genes found on jellyfish *Rhopilema esculentum* chromosome 2 (RES2), amphioxus *Branchiostoma floridae* chromosome 8 (BFL8), and sponge chromosomes CLA16 (of a cladorhizid demosponge) and EMU16 (of *Ephydatia muelleri*)). Note that, by our definition, two different conserved synteny groups can coexist on the same chromosome in some species. For example, amphioxus chromosome BFL5 is seen to be a combination of BCnS ALGs Ea and Eb, which are found on distinct chromosomes in other species.

Our results extend previous findings[12] of BCnS ALGs by incorporating a new chromosome-scale genome sequence of a recently discovered bioluminescent deep-sea cladorhizid demosponge[36] (Fig. 1c, Supplementary Information 2 and Extended Data Fig. 1c–i) complementing the spongillid demosponge *E. muelleri*[37]. Although the cladorhizid and spongillid lineages diverged approximately 450 million years ago[38], chromosomes of the two demosponges correspond simply with each other (Fig. 1d and Extended Data Fig. 2f–h) and with bilaterian and cnidarian chromosomes (Fig. 1d and

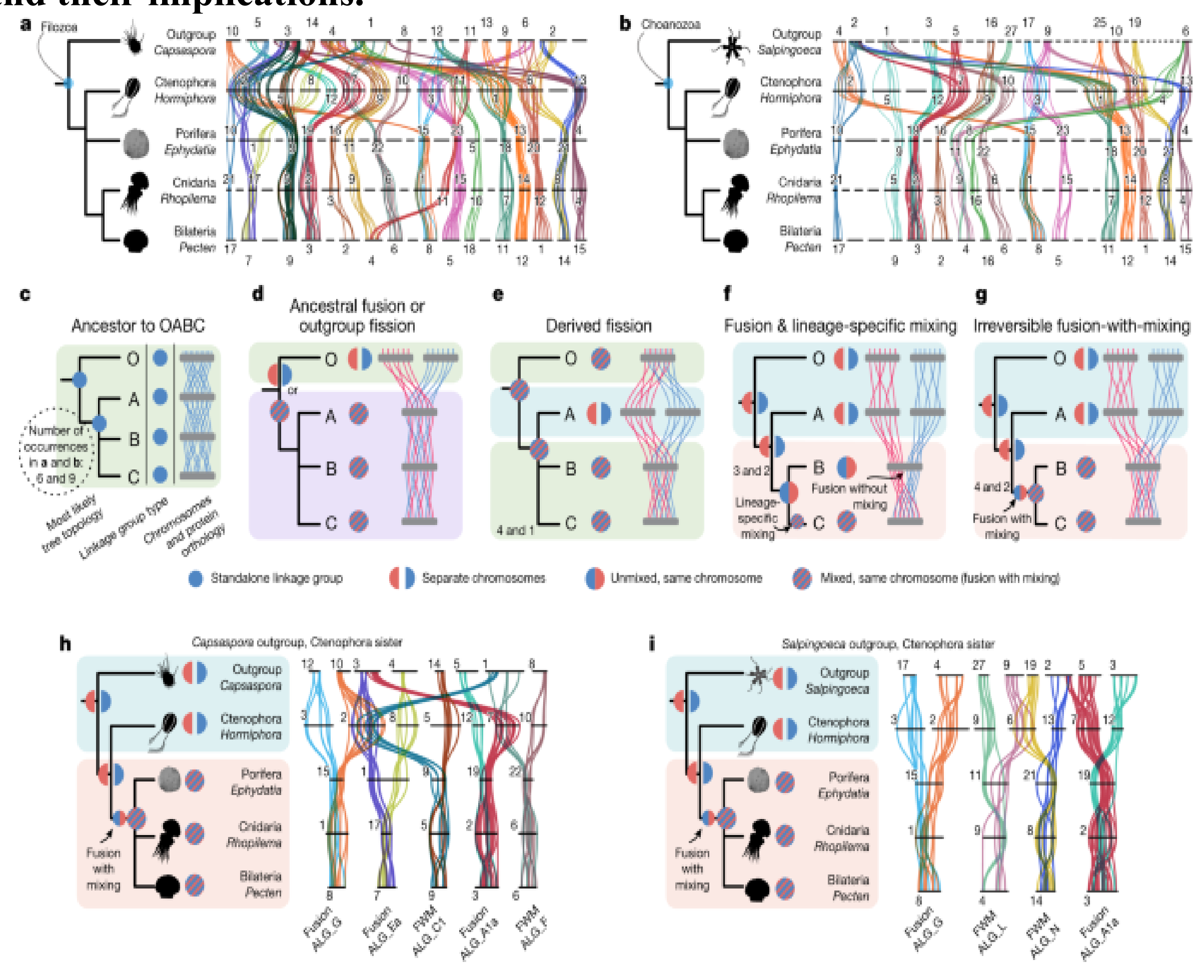
Extended Data Fig. 2i–k), consistent with the previously described genome tectonic schema[12]. Further comparisons with other recently released chromosome-scale demosponge genome sequences[39] confirm the high degree of conserved synteny in this group, but show that one of the rearrangements that we found in the cladorhizid genome is the result of a fission in that lineage (ALG_H; Extended Data Fig. 2f–k). We also sequenced the genome of a previously undescribed hexactinellid (glass) sponge (Extended Data Fig. 1j–n and Supplementary Information 2), but found it to be considerably rearranged. Despite many lineage-specific genomic changes in glass sponges, relicts of 10 out of 29 BCnS ALGs are detectable (Extended Data Fig. 3). Owing to the high degree of rearrangement, we do not consider hexactinellid genomes further.

However, in contrast to demosponges, genomic comparisons between the cydippid ctenophore *Hormiphora californensis*[40] and other metazoans reveal patterns of both conserved and altered synteny (Fig. 1d and Extended Data Fig. 2b–e). For example, whereas the BCnS group ALG_Ea is localized to a single ctenophore chromosome (*H. californensis* chromosome 8 (HCA8)), the BCnS synteny group ALG_A1a (comprising genes found on amphioxus chromosome BFL1, jellyfish chromosome RES2, and sponge chromosomes EMU1 and CLA15) is partitioned across two ctenophore chromosomes (HCA12 and HCA7). To test whether the observed patterns of ctenophore synteny are unique to the *H. californensis* lineage or common across ctenophores, we assembled and analysed the genome of the recently redescribed[41] lobate comb jelly *Bolinopsis microptera* (Fig 1b; the assembly is reported in the Methods, Supplementary Information 1, Extended Data Fig. 1a,b and Supplementary Table 1.1–1.4). Despite the 160–260-million-year divergence between lobate and cydippid ctenophores[10] their $n = 13$ chromosomes show one-to-one correspondence (without gene order conservation) (Fig. 1d and Extended Data Fig. 2a). This finding implies that a common $n = 13$ karyotype is ancestral for the *Hormiphora–Bolinopsis* crown group, and that cross-metazoan patterns of synteny shown in Fig. 1d are general.

Interpreting the differences in synteny between ctenophores and other animals depends on the ancestral metazoan state (Fig. 2c–g, Extended Data Fig. 4 and Supplementary Information 4). If BCnS syntenies are ancestral to

all metazoans, then the partitioned syntenies observed in ctenophores would have arisen by rearrangements that split the ancestral chromosomes in the ctenophore lineage (syntenic autapomorphies; Fig. 2e) and would therefore be uninformative for discriminating between the ctenophore-sister and sponge-sister hypotheses. Alternatively, if the patterns of synteny found in ctenophores are ancestral to animals, the derived syntenies shared by BCnS to the exclusion of ctenophores could have arisen by fusion on the BCnS stem lineage, which would represent syntenic synapomorphies (Fig. 2f,g). In this case, ctenophores would be excluded from the BCnS clade and established as the sister clade of all other extant metazoans. Note that the extensive conservation of synteny between sponges, bilaterians, and cnidarians[12] confirmed here makes it improbable that ctenophores could share syntenies with cnidarians and bilaterians to the exclusion of sponges and, indeed, we did not find any such cases in analyses described below.

Fig. 2: Patterns of conserved synteny between animals and outgroups, and their implications.



a,**b**, Conserved linkages between the chromosomes of animals and two non-animal outgroups ($\alpha \leq 0.05$, permutation test one-sided false-discovery rate). **a**, The filasterean amoeba *C. owczarzaki*. **b**, The choanoflagellate *S. rosetta*. Each synteny group (conserved between metazoans and an outgroup) was assigned a distinct colour (different from Fig. 1). **c**–**g**, Schematics showing phylogenetic information (ancestor to OABC (**c**); ancestral fusion or outgroup fission (**d**); derived fusion (**e**); fusion and lineage-specific mixing (**f**); and irreversible fusion with mixing (**g**)) conveyed by patterns of conserved synteny based on a quartet analysis. Node O designates the outgroup, and nodes A, B and C are ingroups of which the phylogenetic branching is to be determined. The thin red and blue lines in **d**–**g** represent genes of distinct synteny groups on different chromosomes in at least one species. Changes in syntenic characters are indicated schematically on parsimonious phylogenies on the left of each diagram. **d**,**e**, Single-species differences are phylogenetically uninformative. **f**,**g**, Shared chromosomal distributions between outgroup O and one of the ingroups (labelled taxon A) imply that the other two ingroups (taxa B and C) are related by fusion of ancestral synteny groups. **f** and **g** differ in whether all fusions have subsequently mixed. Fusion with mixing (**g**) is the strongest phylogenetic character because it represents an irreversible change, as discussed in the main text. **h**,**i**, Subsets of the synteny groups shown in **a** and **b** (*Capsaspora* (**h**) and *Salpingoeca* (**i**)) that match the phylogenetically informative patterns indicated in **f** and **g**. In all such cases, ctenophore syntenies match the outgroup and sponges share fusions with bilaterians and cnidarians. Note that groups A1a and G are found in both outgroups. We did not observe any cases in which sponge syntenies match the outgroup to the exclusion of ctenophores, bilaterians and cnidarians.

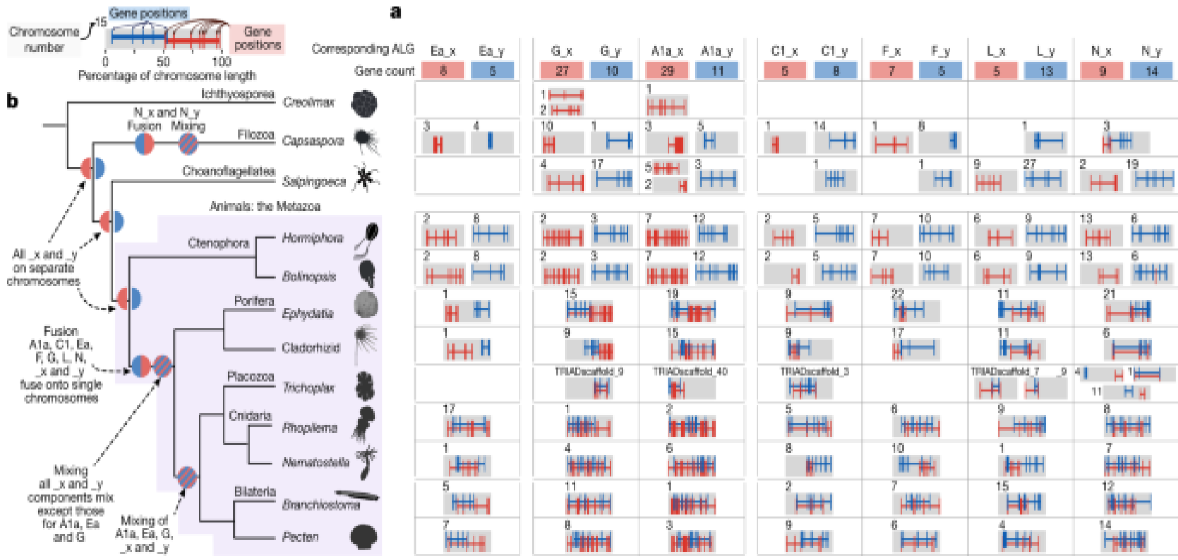
To provide outgroups for inferring ancestral metazoan syntenies, we assembled chromosome-scale sequences of representatives of three unicellular lineages closest to animals (collectively, outgroups): the choanoflagellate *Salpingoeca rosetta* (chromosome number, $n = 36$), the filasterean amoeba *Capsaspora owczarzaki* (chromosome number, $n = 16$) and the ichthyosporean *Creolimax fragrantissima* (chromosome number, $n = 26$). Chromosome-scale sequences and karyotypes were obtained by integrating previously reported subchromosomal draft sequences[42,43,44] with

new chromatin conformation data (Methods, Supplementary Information 3 and Extended Data Fig. 5).

Chromosomal comparisons across animal and non-animal outgroup genomes revealed conserved ancestral metazoan synteny groups ranging in size from 5 to 29 genes, totalling 291 genes (out of 2,474 outgroup-metazoan orthology groups; Methods, Figs. 2 and 3, Extended Data Tables 1 and 2, Extended Data Figs. 5–8, Supplementary Information 5–9 and Supplementary Data 2). This finding extends previous observations based on subchromosomal assemblies of non-animal species[12]. Each such ancestral metazoan synteny group is a collection of genes of which the orthologues are consistently linked on single chromosomes in diverse metazoans and at least one outgroup (Fig. 2h–i and Extended Data Table 2). In contrast to the readily detected conserved syntenies among sponges, cnidarians and bilaterians, conserved syntenies involving ctenophores and non-animal outgroups are not visually evident in pairwise comparisons with other animals (Extended Data Figs. 2 and 6) but are statistically supported in multispecies comparisons (Methods and Supplementary Information 4 and 11). On the basis of permutation tests, the false-discovery rate of a conserved group of five linked genes in a four-species comparison is $\alpha \leq 0.0003$, and groups of eight or more linked genes never occurred in ten million permutations (Supplementary Information 8 and Extended Data Table 2). To maximize coverage of lineages relevant for the branching order of sponges and ctenophores, we considered orthologous genes across quartets of the form {outgroup, sponge, ctenophore, cnidarian/bilaterian}, which does not presuppose either the ctenophore-sister or sponge-sister hypothesis (Fig. 3 and Supplementary Information 4, 8 and 9). The extensive conservation of synteny across BCnS and within ctenophores makes our analysis insensitive to which genomes are used to represent these major metazoan clades (Supplementary Information 8). Here we used the scallop *Pecten maximus*[45], the fire jellyfish *R. esculentum*[46], the freshwater sponge *E. muelleri*[37] and the ctenophore *H. californensis* to represent the bilaterian, cnidarian, sponge and ctenophore genomes (Methods), although our findings do not depend on these choices (Figs. 2 and 3). We used two different methods for identifying orthologues—a simple mutual-best-hits method and an alternative orthologue-clustering approach (OrthoFinder[47];

Supplementary Information 10 and 11), and obtained comparable results using both approaches (Extended Data Fig. 9).

Fig. 3: Phylogenetically informative syntenies support ctenophores as the sister clade to other animals.



a, The rows represent the species considered in our analyses. The top three rows are non-metazoan outgroups. Columns show pairs of phylogenetically informative and significantly large metazoan syntenies ($\alpha \leq 0.05$, permutation test one-sided false-discovery rate), labelled according to their BCnS names (A1a, C1 and so on) with the suffix _x or _y denoting subgroups shared across metazoans. The number of genes participating in each metazoan synteny group is indicated in red and blue rectangles at the top of each column. Only genes with defined orthologues in outgroups are shown here. Extended Data Fig. 8 shows a larger set of genes requiring only metazoan orthologues. Inset: the convention for representing gene distributions on chromosomes (top left). The grey rectangles represent chromosomes (or large scaffolds in the placozoan *Trichoplax*). The chromosome number or scaffold name is located above or to the left of the grey rectangle. Red and blue vertical hashes represent the relative position of genes participating in phylogenetically informative pairs of metazoan synteny groups. **b**, The most parsimonious phylogeny according to the logic of Fig. 2c–g (Extended Data Fig. 4 and Supplementary Information 4), the results of ref. [12] and the accepted monophyly of demosponges.

Although choanoflagellates are considered to be the closest living relatives of animals[48,49], we found that the more distantly related filasterean *Capsaspora* shares 29 conserved synteny groups with metazoans, compared to 20 between the choanoflagellate *Salpingoeca* and metazoans (Extended Data Figs. 6 and 7), perhaps indicating more rapid interchromosomal rearrangement in the *Salpingoeca* lineage. The even more distantly related ichthyosporean *Creolimax* still retains eight conserved synteny groups with metazoans. Although we considered each outgroup-plus-metazoan comparison separately, we found widespread overlap between the ancient synteny groups defined independently by comparison with *Capsaspora* and *Salpingoeca*. In total, our analysis defined 31 ancestral metazoan synteny groups that are traceable to the last common ancestor of Metazoa and shared by one or both of *Capsaspora* and *Salpingoeca* (Extended Data Table 1). The extensive conservation of synteny within BCnS implies that the ancestral metazoan synteny groups correspond to subsets of the BCnS groups, and we name them using the BCnS notation with the suffixes _x and _y. If we relax the condition that an outgroup gene must be present, more metazoan genes can be added to these ancestral metazoan syntenic units (Extended Data Fig. 8).

Conservation of synteny between animals and their unicellular relatives may at first seem surprising, as these lineages diverged more than 800 million years ago[50]. Within animals, it has been estimated that ongoing small-scale translocations between chromosomes typically transfer 1% of genes to a different chromosome every ~40 million years[12]. The limited residual conservation of synteny between animals and close unicellular relatives suggests that small-scale translocations have occurred at similarly low rates along both choanoflagellate and filasterean lineages. The more extensive conservation observed between animals and *Capsaspora* versus *Salpingoeca* may be due to variations in this rate or differences in other chromosomal rearrangements over deep time. The *Capsaspora* karyotype is predominantly metacentric and, notably, we find that 11 of the 29 ancient synteny groups found in *Capsaspora* are concentrated on single chromosome arms, rather than dispersed across whole *Capsaspora* chromosomes, based on estimates of centromere position using chromatin conformation contacts. This raises the possibility that *Capsaspora* chromosome arms preserve ancient filozoan units and suggests further attention to the chromosome biology of non-

metazoan relatives. We found no significant functional associations of anciently linked groups of genes (Supplementary Information 12 and Supplementary Data 3), consistent with a general slow rate of synteny loss due to the infertility of translocation heterozygotes[28,31], which allows only small-scale interchromosomal translocation[32].

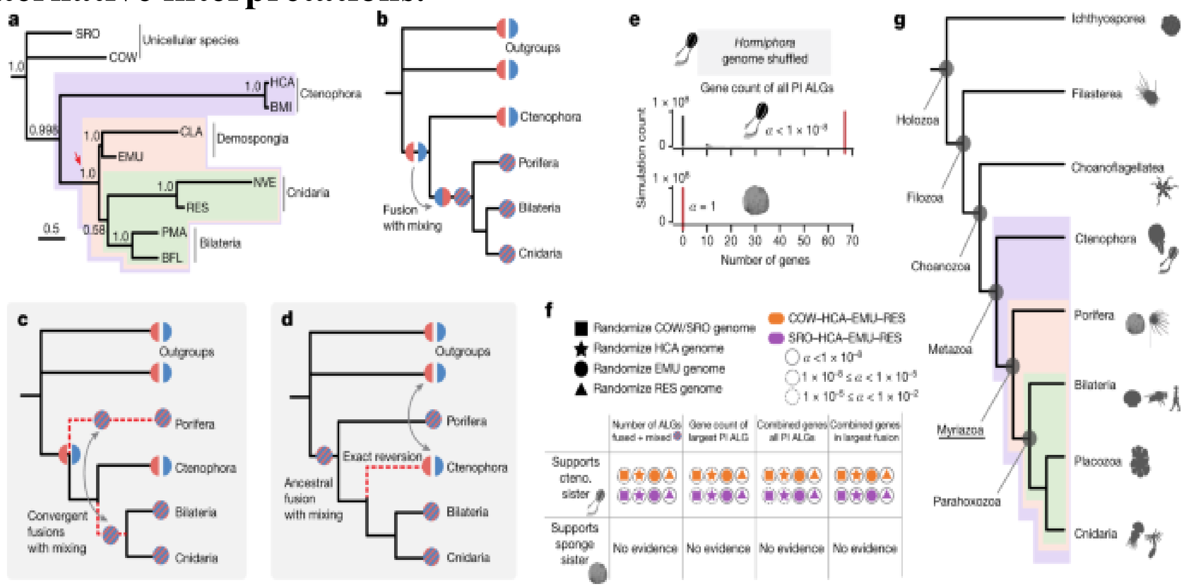
With conserved ancestral metazoan syntenies in hand, we tested the ctenophore-sister versus sponge-sister hypotheses by identifying shared, derived syntenic characters using standard phylogenetic methods. As noted above, two or more metazoan synteny groups can co-occur on the same chromosome in one or more genomes, corresponding to ancient fusions (that is, translocations[51]) (Fig. 2h,i). As only shared derived characters are phylogenetically informative, changes that are unique to a single lineage can be disregarded (Fig. 2e and Extended Data Fig. 4b,d,e,g). There are two different types of chromosomal fusions between two ancestrally linked groups of genes: without mixing (Fig. 2f) or with subsequent intermixing (Fig. 2g and Supplementary Data 4 and 5). Fusion-without-mixing is potentially reversible, as observed in Robertsonian fusions and fissions involving whole chromosome arms[51]. However, in the fusion-with-mixing case, reversion is extremely unlikely, comparable to the spontaneous reduction of entropy after mixing of two fluids[12].

We encoded the state of each potential fusion into a phylogenetic character matrix as 0 (no fusion, that is, ALGs found on separate chromosomes), 1 (fused but unmixed) or 2 (fused and mixed). The mixed/unmixed status of a fusion was determined on the basis of the likelihood of the observed gene arrangement under a model of random rearrangement (that is, entropy of mixing of the two fused groups) (Methods, Supplementary Information 13 and Supplementary Data 4 and 5). The same fusion character states were obtained using orthology defined by mutual-best-hits or OrthoFinder. We then applied the machinery of Bayesian phylogenetics[52] to this character matrix, using asymmetric transition probabilities to reflect the highly improbable unmixing transition (Methods and Supplementary Data 6).

Bayesian phylogenetic analyses of the fusion character matrix strongly support the ctenophore-sister topology (Fig. 4 and Supplementary Information 14). The same conclusion is clear from direct examination of

the fusions identified in our data. Specifically, there are seven derived fusions shared by bilaterians, cnidarians and sponges to the exclusion of ctenophores (Extended Data Table 1). Of these seven derived fusions, four are accompanied by mixing of genes from two different ancestral chromosomes—a process that is essentially irreversible (Figs. 2e and 4b–d); the other three are mixed only in bilaterians and cnidarians (Fig. 3).

Fig. 4: Phylogenetic analysis of patterns of conserved synteny and alternative interpretations.



a, Bayesian phylogenetic analysis of conserved syntenies supports monophyly of the group comprising demosponges, Cnidaria and Bilateria, to the exclusion of Ctenophora, with high posterior probability (red arrow: 1.0, 100,000 generations with 25% burn-in). Bayesian analysis was run on both constrained (per-phylum constrained tree shown) and unconstrained tree topologies (Supplementary Information 14 and Supplementary Data 6). This panel corresponds to Supplementary Fig. 14.1c. **b**, Character transitions involving ALGs C1, F, L and N in Fig. 3 are most parsimoniously interpreted as fusion with mixing on the myriazoan stem after divergence from ctenophores, which retain the ancestral metazoan state as inferred from outgroup comparisons. **c**,**d**, To interpret the observed patterns under the alternative sponge-sister hypothesis would require unlikely convergent chromosomal changes (either convergent fusions (**c**) or exact unmixing and fissions to the ancestral state (**d**)) that were not seen in our genomes. **e**, The number of genes in the genome-shuffling simulations ($n = 1 \times 10^8$) that

support the ctenophore-sister (upper) or sponge-sister (lower) hypothesis. For the ctenophore *Hormiphora*, the number of fusion-with-mixing events is significantly higher in the observed genomes (vertical red bars) than in the *Hormiphora* genome-shuffling simulations (vertical grey histogram bars). Significance is shown as the one-sided false-discovery rate, α, of a genome-shuffling permutation test. There were no groups of genes that supported the sponge-sister hypothesis in the real genomes, and none occurred in the genome-shuffling simulations. **f**, Additional statistical measures also support only the ctenophore (cteno.)-sister hypothesis in genome-shuffling simulations of *Hormiphora*, *Capsaspora* (COW), *Salpingoeca* (SRO), *Ephydatia* (EMU) and *Rhopilema*. PI ALG, phylogenetically informative linkage groups. The shape indicates the treatment; the colour indicates the outgroup. The full figure is shown in Extended Data Fig. 10. **g**, Summary of phylogenetic relationships among animals and close outgroups including syntenic characters. Myriazoa (underlined) is the name proposed for the clade containing extant animals, except Ctenophora. Outgroup topology follows ref. [49].

We reject the alternative sponge-sister hypothesis as it would require either (1) multiple convergent fusions (that is, involving the same groups of genes) in both the sponge and bilaterian-cnidarian lineages (Fig. 4c and Supplementary Information 15) or (2) the precise reversal of multiple fusions-with-mixings in the ctenophore lineage to match the original patterns found in the ancestral metazoan lineage (Fig. 4d). The extreme unlikeliness of recovering the observed syntenic patterns by chance is shown by simulations in which we permuted the configuration of the genes in each of the *C. owczarzaki*, *S. rosetta*, *H. californensis*, *E. muelleri* and *R. esculentum* genomes (Supplementary Information 15 and Extended Data Fig. 10). Across one hundred million randomized *Hormiphora* genomes, we never found syntenic signals comparable to those observed with the actual genome, indicating that syntenic support for ctenophore-sister is unlikely to have arisen by chance (Fig. 4e). We also note the complete absence of syntenic synapomorphies of a hypothetical ctenophore–bilaterian–cnidarian clade that excludes sponges, both in the actual data and in genome-shuffling simulations (Fig. 4f, Extended Data Fig. 10 and Extended Data Tables 1 and 2). This lack of homoplasy allows for a simple interpretation of the results (Fig. 4).

Conclusions

Support for the ctenophore-sister hypothesis

Our findings provide strong support for the ctenophore-sister scenario and reject the sponge-sister hypothesis. Although we encoded syntenic states as a character matrix and analysed it using a Bayesian phylogenetic framework (Fig. 4a and Supplementary Fig. 14.1), the cladistic logic supporting our conclusions is easily appreciated, as emphasized above (Fig. 2, Extended Data Fig. 4 and Supplementary Information 4). Previous phylogenetic analyses of sequence-based characters have not resolved the sponge-sister versus ctenophore-sister hypotheses because the phylogenetic signal is weak and distributed across thousands of individual amino acid positions that are often saturated or subject to confounding evolutionary forces[11]. By contrast, the synteny-based characters that support ctenophores as sister to other animals in our analysis are clear: sponges, bilaterians, and cnidarians share multiple irreversible changes in synteny to the exclusion of ctenophores (BCnS syntenic synapomorphies) (Figs. 2 and 3). Support for the ctenophore-sister hypothesis is directly testable by future genome sequencing, as it is a strong prediction of our model that all bilaterian, placozoan, cnidarian or sponge genomes should share the four fusion-with-mixing syntenic synapomorphies shown in Figs. 2 and 3 and, to a lesser extent, the three fusion-without-mixing events (pending considerations of sponge monophyly; Supplementary Information 7.2.6). The placement of ctenophores as sister to other animals also rejects the old notion of a Coelenterata clade that would unite ctenophores with cnidarians[53].

Myriazoans

The clade containing all sponges, bilaterians, cnidarians and placozoans is diverse, accounting for all living animals other than ctenophores. In recognition of this morphological diversity, we propose that this clade be called Myriazoa, from the Greek myria (extremely great in number) and zoa (animals) (Fig. 4g). While Myriazoa is supported by shared derived chromosomal fusions, there are currently no obvious morphological characters that unite them. The name Benthozoa was proposed for this

clade[23] on the basis of the inference of a pelagic ancestral metazoan and a derived benthic adult ancestor of the clade sister to ctenophores, but a benthic life history stage may not be a shared derived feature of this clade. In particular, it would be just as parsimonious for the ancestor of Metazoa to have had a benthic stage, and for most ctenophores to have lost it. We therefore prefer to avoid any assumption of the ancestral life history strategy in referring to the clade.

Parahoxozoans, sponges and placozoans

A clade grouping bilaterians, placozoans and cnidarians to the exclusion of sponges and ctenophores[54] has been recovered in multiple phylogenetic studies[8,10] and is now called Parahoxozoa on the basis of the shared presence of Hox/ParaHox-class genes[14]. Parahoxozoa is supported in our analysis by the disposition of the ancestral myriazoan linkage groups Ea and G, which are each partitioned across two chromosomes in non-metazoan outgroups and ctenophores. The pre-myriazoan partitions of Ea and G are fused in demosponges and parahoxozoans, but are mixed only in parahoxozoans, providing a candidate parahoxozoan synapomorphy. The most parsimonious interpretation is that fusions forming Ea and G occurred without mixing on the myriazoan stem, a state that is preserved in demosponges, but that mixing occurred on the parahoxozoan stem lineage so that the mixed state is shared by all bilaterians, cnidarians and placozoans. However, a detailed understanding of the history of Ea and G linkages in sponges will require chromosome-scale genome sequences from other sponge classes beyond demosponges and lyssacinosid glass sponges (Supplementary Information 7.2.2, 7.2.5 and 7.2.6). If sponges are monophyletic (as supported by recent phylogenomic studies[8,10,55]), then the four fusions-with-mixing that are found in demosponges and parahoxozoans must be shared by all sponges. However, if one or more sponge classes branched before the split between the demosponge and parahoxozoan lineages, it is possible that the descendants of the early-branching sponges might not possess one or more of these myriazoan fusions-with-mixing.

Although the subchromosomal assemblies currently available for *Trichoplax* preclude its full integration into the present analysis, Fig. 3 shows that placozoans share the diagnostic myriazoan fusion-with-mixing characters

related to ALG_C1 and the two bilaterian–cnidarian fusions-with-mixing related to ALGs A1a and G. The placozoa-sister-to-other-animals hypothesis[56] is rejected by the placement of placozoans within Myriazoa using synteny. It is therefore a strong prediction of our overall approach that chromosome-scale assemblies of placozoans will show that they share the fusions and mixing events that define Myriazoa. Furthermore, we previously showed that cnidarians and placozoans are united as sister lineages to the exclusion of bilaterians and sponges based on the mixing of genes from ALG_Ea and ALG_F found on cnidarian chromosomes and placozoan scaffolds[57], consistent with recent gene trees[55]. These characters do not appear in the present analysis owing to the stringent requirement that syntenies considered here are also preserved in outgroups to Metazoa. If placozoans are nested within Parahoxozoa, homologies between the mouth, gut and nervous systems of cnidarians and bilaterians imply that placozoans are secondarily flattened and have lost an ancestral nervous system, rather than representing the ancestral parahoxozoan state.

Implications for early animal evolution

Finally, we consider implications of the ctenophore-sister hypothesis for early animal evolution[1,2]. Comparisons among diverse genomes have identified numerous genes that are present in myriazoans but are absent in ctenophores[1,5,58]. Under the ctenophore-sister scenario, these are most parsimoniously interpreted as arising on the myriazoan stem after the divergence of ctenophores[1,5,59], and include genes associated, in bilaterians and cnidarians, with neuronal function[1,5,59], development[58] and cell adhesion[60]. However, as gene loss is common throughout animal evolution[61], it is also possible that some of these genes were present in the ancestral metazoan but lost in ctenophores. Similarly, some genes are present in ctenophores and parahoxozoans but absent in sponges[58], and these must be interpreted as gene losses on the sponge lineage.

Perhaps the most intriguing suite of metazoan characters pertain to neuromuscular systems, which are present in varying complexity in ctenophores, bilaterians, and cnidarians but are absent in sponges[3,6,59]. In sponge-sister scenarios, these characters are interpreted as being primitively absent, arising after the divergence of sponges on the stem lineage leading to

other animals. However, in the ctenophore-sister scenario supported here by deeply conserved syntenies, there are two possible alternatives explaining the evolution of neurons: either complex neural systems arose more than once[3,59,62] but were elaborated differently in ctenophores, cnidarians and bilaterians[3,4,59], or neuronal cell types were present in the metazoan ancestor but were lost in the sponge lineage[4,9,63].

Sponge-sister and ctenophore-sister hypotheses are sometimes erroneously interpreted as suggesting that the most recent common ancestor of animals was sponge-like or ctenophore-like. We must be mindful, however, that the living representatives of sponges, ctenophores, bilaterians and placozoans may be poor surrogates for the earliest members of each stem-lineage, as the crown group of each clade arose hundreds of millions of years after their divergence from each other, let alone from the common metazoan ancestor[2]. Although living sponges are often defined by the cellular, morphological and developmental characters that they lack relative to other animals, they are complex animals in their own right, successfully adapted to a unique benthic filter-feeding lifestyle[13]. Consistent with a neuron-bearing metazoan ancestor, sponges possess secretory cell types[15] and extensive molecular components associated with presynaptic function that could be derived from a primitive neurosecretory cell. Conversely, the elaborate and divergent nervous systems of living ctenophores, bilaterians, and cnidarians do not represent the stem ancestors of these groups, which would have had very different lifestyles in the Ediacaran. The nervous systems of living ctenophores, cnidarians, and bilaterians each have unique properties[6,19,59], and could represent divergent evolution from a simpler neuron-bearing common ancestor. With the ctenophore-sister topology in hand, reconstructing the characters of this metazoan ancestor will require an improved understanding of molecular, cellular and system homologies and specializations across the full range of animal diversity.

Methods

A full description of the methods is provided in the Supplementary Information.

Unicellular outgroup species genome scaffolding

Chromatin conformation capture (Hi-C) libraries were generated[64] from frozen cell cultures obtained directly from the American Type Culture Collection. The cultures used were of the species *C. owczarzaki* (ATCC, 30864), *C. fragrantissima* (ATCC, PRA-284) and *S. rosetta* (ATCC, PRA-366). The strains used were the same as those sequenced in the original genome assembly projects for each species[42,43,44]. The Hi-C libraries were sequenced at a depth of over 500× for each species on the Illumina NovaSeq 6000 system at MedGenome.

Previously published draft genome assemblies[42,44] were scaffolded to chromosome-scale using a combination of HiRise (v.Aug2019)[65] and SALSA2 (v.2.3)[66]. The genomes were manually curated using PretextView v.0.2.4 (https://github.com/wtsi-hpag/PretextView), HiGlass v.1.10.0104[67], Juicebox Assembly Tools (GitHub Commit 46c7ed1105)[68], the Juicebox visualization system (v.1.11.08106)[69] and artisanal (https://bitbucket.org/bredeson/artisanal/src). For *C. owczarzaki*, we used the most recent 'v4' assembly as input for scaffolding[70]. The Hi-C data were used as evidence to remove several megabases of the original *C. fragrantissima* assembly that, after further analysis, appeared to be fungal contaminants. We identified the general location of the centromeres in *C. fragrantissima* and *C. owczarzaki* using the Hi-C data as described in Supplementary Information 3.

Sponge and ctenophore genome assembly

Samples of *B. microptera*[41] were collected in Monterey Bay, California (36.63° N, 121.90° W) from surface waters and were reared to an F_3 population at the Monterey Bay Aquarium, from which one adult was sequenced. One individual cladorhizid sponge[36] was collected off the coast of Big Sur, California (35.49° N, 124° W) from the seafloor at 3,975 m. One hexactinellid 'tulip' sponge (HEX) was collected near Southern California (34.57° N, 122.56° W) from the seafloor at 3,852 m. This species of ctenophore, and presumably these species of sponges, are hermaphroditic. Sponge and ctenophore samples were collected under the State of California

Department of Fish and Wildlife collecting permits SC-2026 (*Bolinopsis*) and SC-4029 (sponges).

DNA and RNA were isolated from these species to generate Pacific Biosciences (PacBio) CLR WGS, HiFi WGS libraries or PacBio Iso-Seq libraries at the Brigham Young University DNA Sequencing Center. These libraries were sequenced on the PacBio Sequel II system. Illumina WGS libraries, Chicago libraries and Hi-C libraries were generated at UC Santa Cruz and sequenced at MedGenome on the Illumina HiSeq X system. PacBio WGS library coverage was over 70× for all three species, and Hi-C coverage was over 190× for all three species. Genome sizes were estimated using jellyfish (v.2.2.10)[71], then using the resulting spectrum in GenomeScope (v.2)[72].

The genome of *B. microptera* was assembled using wtdbg (v.2.4)[73], and the sponge genomes were assembled using hifiasm (v.0.16.1-r375)[74]. Hi-C reads were mapped using bwa mem (v.0.7.17)[75], processed using pairtools (v.0.3.0)[76], pairix (v.0.3.7; https://github.com/4dn-dcic/pairix) and Cooler (v.0.8.10)[77], and scaffolding was performed using HiRise (v.Aug2019)[65]. In *B. microptera*, gaps were closed using TGS-Gapcloser (v.1.1.1)[78], haplotigs were removed using Purge Haplotigs (v.1.0.4)[79] and the assembly was polished using Illumina WGS reads and pilon (v.1.23)[80]. In both the sponge and *B. microptera* genomes, bacterial scaffolds were removed using Diamond (v.0.9.24)[81] and Blobtools (v.1.0)[82]. The genomes were manually curated with Hi-C data as described above. The haplotypes of the hifiasm-based assemblies were compared to one another using D-Genies (v.1.4.0)[83].

Genome annotations

The unicellular outgroup genome assemblies were annotated by mapping their transcripts from the original assemblies to the Hi-C scaffolded assemblies using minimap2 (v.2.23)[84]. To clarify demosponge macrosyntenic relationships, we produced putative *Ephydatia* protein coordinates in the cladorhizid sponge using tblastn (v.2.10.0+)[85]. To annotate the hexactinellid sponge genome, we mapped the proteins of closely-related hexactinellid species[86,87] using miniprot (v.0.2)[88]

(Supplementary Information 2.1.5). The *Bolinopsis* genome was annotated using BRAKER (v.2.14)[89] supplied with evidence from RNA-seq reads mapped with STAR (v.2.7.1a)[90] and minimap2 (v.2.23)[84], Iso-Seq reads processed with lima (v.2.2.0; https://github.com/PacificBiosciences/barcoding) and isoseq3 (v.3.4.0; https://github.com/PacificBiosciences/IsoSeq) then mapped with minimap2 (v.2.23)[84], and protein orthology identified using ProtHint (v.2.6.0)[91] from ctenophore transcriptomes[92,93,94] assembled with Trinity (v.2.5.1)[95] and translated using TransDecoder (v.5.5; https://github.com/TransDecoder/TransDecoder). We assessed genome sequence and protein datasets using BUSCO (v.5)[96].

Orthologue Inference

Orthologues were inferred between species by finding reciprocal-best BLASTp[97] hits between the proteins in the genomes, or with OrthoFinder (v.2.3.7)[98]. The reciprocal-best BLASTp hits were used to identify macrosyntenic chromosomes between species by performing Bonferroni-corrected one-sided Fisher's exact tests[57]. To determine the provenance of the ALG_H in sponges, the genomes of *Chondrosia* and *Petrosia*[39,99], *Oopsacas*[86], CLA and HEX were compared using the odp software suite.

Orthologues shared between three, four or more species were selected by finding groups of proteins that were *n*-way reciprocal best BLASTp hits. In this conservative method, each orthogroup has a single protein from each of the *n* species. We performed this analysis for three-way and four-way comparisons of combinations of the species CFR, COW, SRO, HCA, EMU, CLA, RES, BFL, NVE and *P. maximus*.

Gene linkage group identification

Orthologues from three-way or four-way reciprocal-best BLASTp searches were grouped by the chromosomes on which the genes occurred in the *n* species. To identify which sets of orthologues were larger than expected by random chance, we shuffled the genome coordinates of the *n* species and measured the frequency of finding sets of orthologues of size *k* on the same

chromosomes in the n species. By performing this for 10 million iterations, we calculated the false-discovery rate (α) of finding an orthologue set of size k given the n input genomes.

Combined unicellular outgroup analysis

Sets of orthologues with a false-discovery rate of less than 0.05 were retained from the four-way reciprocal best hit searches of COW–HCA–EMU–RES, CFR–HCA–EMU–RES and SRO–HCA–EMU–RES. The remaining orthogroups were joined based on gene identity in HCA–EMU–RES, such that each orthologue contained a protein from at least one of the unicellular outgroup species. This yielded 291 sets of orthologues.

Identification of orthologues in other species

For each of the 291 orthologues, we aligned the proteins using MAFFT (v.7.310)[100], built a hidden Markov Model using hmmbuild in hmmer (v.3.3.2)[101], then found the best match using hmmsearch in the proteins of the genomes of other species, including the ctenophore *B. microptera*, the cladorhizid sponge, *T. adhaerens*[102], *H. vulgaris*[12], *N. vectensis*[103], *B. floridae*[57], *P. maximus*[45] and *E. muelleri*[37]. To test for Gene Ontology enrichment of the sets of orthogroups using PANTHER (v.17)[104], we also searched for the orthologues in *Homo sapiens*[105].

Mixing analysis

To test whether the _x and _y gene sets present on single chromosomes were well-mixed, we used a metric that counts the number of transitions between a gene in _x to a gene in _y and vice versa. To provide an objective measure of mixing, we computed the α value (false-discovery rate) that the two sets of genes are unmixed by building a distribution of mixing scores from randomly sorted groups of the same size of the _x and _y groups in question. We consider $\alpha < 0.05$ to be unmixed.

Simulations testing the ctenophore-sister and sponge-sister hypotheses

We applied this methodology to test whether the findings supporting the ctenophore-sister hypothesis were due to the arrangement of any of the observed genomes, implemented as part of the odp software suite. For both the SRO–HCA–EMU–RES and COW–HCA–EMU–RES four-way reciprocal best hit results, we performed four analyses. One analysis shuffles the genome chromosome labels of one species 100 million times. Each time the genome chromosome labels are shuffled, we perform the gene linkage group identification analysis described above, and measure the quantity and size of gene linkage groups that support either the ctenophore-sister or sponge-sister hypothesis. The distribution of these results compared with the observed data of the real genomes is used to estimate the false-discovery rate of finding support for the ctenophore-sister hypothesis or sponge-sister hypothesis. We modelled fusion-with-mixing events in the animal genomes as state transitions, and used RevBayes (v.1.1.1)[106] and MrBayes (v.3.2.7a)[52] to estimate the likelihood of the ctenophore-sister hypothesis, and we used FigTree (v.1.4.4; https://github.com/rambaut/figtree) to visualize the trees.

Software

We implemented a suite of tools for identifying orthologues, plotting syntenic relationships and performing synteny-based phylogenetic analyses using a tool called odp, implemented in snakemake (v.7)[107] for scalability. To confirm the validity of these methods, we applied them to several genome quartets and showed that odp recovers previously identified synapomorphic chromosomal fusion-with-mixing events[12] in bilaterians and cnidarians (Supplementary Information 6).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data presented in this Article are available in public repositories. The sequencing reads are available in the NCBI database under BioProject

accession numbers PRJNA818620, PRJNA818630, PRJNA903214 and PRJNA818537. The genomes for each species are available through the above BioProject accession codes, with the exception of the genomes of *C. fragrantissima*, *C. owczarzaki* and *S. rosetta*, which are available at Dryad (https://doi.org/10.5061/dryad.dncjsxm47). The results shown in the Supplementary Information, when not contained in figures, are also available in the Dryad repository (https://doi.org/10.5061/dryad.dncjsxm47). Publicly available sequencing data and genomes were downloaded from NCBI from BioProject accession numbers PRJNA168, PRJDB8655, PRJNA12874, PRJNA20249, PRJNA20341, PRJEB28334, PRJNA30931, PRJNA31257, PRJNA37927, PRJEB56075, PRJEB56892, PRJNA64405, PRJNA193541, PRJNA193613, PRJNA213480, PRJNA278284, PRJNA281977, PRJNA283290, PRJNA377365, PRJNA396415, PRJNA512552, PRJNA544471, PRJNA576068, PRJNA579531, PRJNA625562, PRJNA667495, PRJNA761294 and PRJNA814716. The *E. muelleri* genome was downloaded from https://spaces.facsci.ualberta.ca/ephybase/.

Code availability

The scripts and software developed for this manuscript are available at Dryad (https://doi.org/10.5061/dryad.dncjsxm47) and Zenodo (https://doi.org/10.5281/zenodo.7857390). Long-term development of odp is available at GitHub (https://github.com/conchoecia/odp).

References

1. Ryan, J. F. et al. The genome of the ctenophore *Mnemiopsis leidyi* and its implications for cell type evolution. *Science* **342**, 1242592 (2013).

2. Halanych, K. M. The ctenophore lineage is older than sponges? That cannot be right! Or can it? *J. Exp. Biol.* **218**, 592–597 (2015).

3. Jékely, G., Paps, J. & Nielsen, C. The phylogenetic position of ctenophores and the origin(s) of nervous systems. *Evodevo* **6**, 1 (2015).

4. Ryan, J. F. & Chiodin, M. Where is my mind? How sponges and placozoans may have lost neural cell types. *Philos. Trans. R. Soc. Lond. B* **370**, 20150059 (2015).

5. Moroz, L. L. et al. The ctenophore genome and the evolutionary origins of neural systems. *Nature* **510**, 109–114 (2014).

6. Burkhardt, P. Ctenophores and the evolutionary origin(s) of neurons. *Trends Neurosci.* **45**, 878–880 (2022).

7. Philippe, H. et al. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).

8. Simion, P. et al. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr. Biol.* **27**, 958–967 (2017).

9. Dunn, C. W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).

10. Whelan, N. V. et al. Ctenophore relationships and their placement as the sister group to all other animals. *Nat. Ecol. Evol.* **1**, 1737–1746 (2017).

11. Li, Y., Shen, X.-X., Evans, B., Dunn, C. W. & Rokas, A. Rooting the animal tree of life. *Mol. Biol. Evol.* **38**, 4322–4333 (2021).

12. Simakov, O. et al. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).

13. Dunn, C. W., Leys, S. P. & Haddock, S. H. D. The hidden biology of sponges and ctenophores. *Trends Ecol. Evol.* **30**, 282–291 (2015).

14. Ryan, J. F. et al. The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* **1**, 9 (2010).

15. Musser, J. M. et al. Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. *Science* **374**, 717–723 (2021).

16. Harbison, G. R. in *The Origins and Relationships of Lower Invertebrates* Systematics Association Special Vol. 28 (eds Morris, S. C. et al.) 78–100 (Clarendon Press, 1985).

17. Tamm, S. L. Formation of the statolith in the ctenophore *Mnemiopsis leidyi*. *Biol. Bull.* **227**, 7–18 (2014).

18. Burton, P. M. Insights from diploblasts; the evolution of mesoderm and muscle. *J. Exp. Zool. B* **310**, 5–14 (2008).

19. Sachkova, M. Y. et al. Neuropeptide repertoire and 3D anatomy of the ctenophore nervous system. *Curr. Biol.* **31**, 5274–5285 (2021).

20. Antcliffe, J. B., Callow, R. H. T. & Brasier, M. D. Giving the early fossil record of sponges a squeeze. *Biol. Rev. Camb. Philos. Soc.* **89**, 972–1004 (2014).

21. O'Brien, L. J. & Caron, J.-B. A new stalked filter-feeder from the middle Cambrian Burgess Shale, British Columbia, Canada. *PLoS ONE* **7**, e29233 (2012).

22. Mah, J. L. & Leys, S. P. Think like a sponge: the genetic signal of sensory cells in sponges. *Dev. Biol.* **431**, 93–100 (2017).

23. Erives, A. & Fritzsch, B. A screen for gene paralogies delineating evolutionary branching order of early metazoa. *G3* **10**, 811–826 (2020).

24. Shen, X.-X., Hittinger, C. T., Rokas, A., Minh, B. Q. & Braun, E. L. Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 0126 (2017).

25. Francis, W. R. & Canfield, D. E. Very few sites can reshape the inferred phylogenetic tree. *PeerJ* **8**, e8865 (2020).

26. Kapli, P., Yang, Z. & Telford, M. J. Phylogenetic tree building in the genomic age. *Nat. Rev. Genet.* **21**, 428–444 (2020).

27. Renwick, J. H. The mapping of human chromosomes. *Annu. Rev. Genet.* **5**, 81–120 (1971).

28. Muller, H. J. in *The New Systematics* (ed. Huxley, J.) 185–268 (Oxford Univ. Press, 1940).

29. Drosophila 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218 (2007).

30. Hane, J. K. et al. A novel mode of chromosomal evolution peculiar to filamentous Ascomycete fungi. *Genome Biol.* **12**, R45 (2011).

31. Wright, S. On the probability of fixation of reciprocal translocations. *Am. Nat.* **75**, 513–522 (1941).

32. Lv, J., Havlak, P. & Putnam, N. H. Constraints on genes shape long-term conservation of macro-synteny in metazoan genomes. *BMC Bioinform.* **12**, S11 (2011).

33. Hillis, D. M. SINEs of the perfect character. *Proc. Natl Acad. Sci. USA* **96**, 9979–9981 (1999).

34. Rokas, A. & Holland, P. W. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* **15**, 454–459 (2000).

35. Okada, N. SINEs: Short interspersed repeated elements of the eukaryotic genome. *Trends Ecol. Evol.* **6**, 358–361 (1991).

36. Martini, S., Schultz, D. T., Lundsten, L. & Haddock, S. H. D. Bioluminescence in an undescribed species of carnivorous sponge (Cladorhizidae) from the deep sea. *Front. Mar. Sci.* **7**, 576476 (2020).

37. Kenny, N. J. et al. Tracing animal genomic evolution with the chromosomal-level assembly of the freshwater sponge *Ephydatia muelleri*. *Nat. Commun.* **11**, 3676 (2020).

38. Schuster, A. et al. Divergence times in demosponges (Porifera): first insights from new mitogenomes and the inclusion of fossils in a birth-death clock model. *BMC Evol. Biol.* **18**, 114 (2018).

39. McKenna, V. et al. The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the tree of life. *Wellcome Open Res.*

6, 254 (2021).

40. Schultz, D. T. et al. A chromosome-scale genome assembly and karyotype of the ctenophore *Hormiphora californensis*. *G3* **11**, jkab302 (2021).

41. Johnson, S. B. et al. Speciation of pelagic zooplankton: invisible boundaries can drive isolation of oceanic ctenophores. *Front. Genet.* **13**, 970314 (2022).

42. Fairclough, S. R. et al. Premetazoan genome evolution and the regulation of cell differentiation in the choanoflagellate *Salpingoeca rosetta*. *Genome Biol.* **14**, R15 (2013).

43. Suga, H. et al. The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* **4**, 2325 (2013).

44. de Mendoza, A., Suga, H., Permanyer, J., Irimia, M. & Ruiz-Trillo, I. Complex transcriptional regulation and independent evolution of fungal-like traits in a relative of animals. *eLife* **4**, e08904 (2015).

45. Kenny, N. J. et al. The gene-rich genome of the scallop *Pecten maximus*. *Gigascience* **9**, giaa037 (2020).

46. Li, Y. et al. Chromosome-level reference genome of the jellyfish *Rhopilema esculentum*. *Gigascience* **9**, giaa036 (2020).

47. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

48. Lang, B. F., O'Kelly, C., Nerad, T., Gray, M. W. & Burger, G. The closest unicellular relatives of animals. *Curr. Biol.* **12**, 1773–1778 (2002).

49. Torruella, G. et al. Phylogenetic relationships within the Opisthokonta based on phylogenomic analyses of conserved single-copy protein domains. *Mol. Biol. Evol.* **29**, 531–544 (2012).

50. Strassert, J. F. H., Irisarri, I., Williams, T. A. & Burki, F. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat. Commun.* **12**, 1879 (2021).

51. Schubert, I. & Lysak, M. A. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet.* **27**, 207–216 (2011).

52. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

53. Hyman, L. H. The phylogeny of the lower metazoa. *Q. Rev. Biol.* **51**, 344–347 (1976).

54. Collins, A. G. Evaluating multiple alternative hypotheses for the origin of Bilateria: an analysis of 18S rRNA molecular evidence. *Proc. Natl Acad. Sci. USA* **95**, 15458–15463 (1998).

55. Laumer, C. E. et al. Revisiting metazoan phylogeny with genomic sampling of all phyla. *Proc. Biol. Sci.* **286**, 20190831 (2019).

56. Schierwater, B. My favorite animal, *Trichoplax adhaerens*. *Bioessays* **27**, 1294–1302 (2005).

57. Simakov, O. et al. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* **4**, 820–830 (2020).

58. Fernández, R. & Gabaldón, T. Gene gain and loss across the metazoan tree of life. *Nat. Ecol. Evol.* **4**, 524–533 (2020).

59. Moroz, L. L. Multiple origins of neurons from secretory cells. *Front. Cell Dev. Biol.* **9**, 669087 (2021).

60. Belahbib, H. et al. New genomic data and analyses challenge the traditional vision of animal epithelium evolution. *BMC Genom.* **19**, 393 (2018).

61. Albalat, R. & Cañestro, C. Evolution by gene loss. *Nat. Rev. Genet.* **17**, 379–391 (2016).

62. Parker, G. H. *The Elementary Nervous System* (J.B. Lippincott Company, 1919).

63. Leys, S. P. Elements of a 'nervous system' in sponges. *J. Exp. Biol.* **218**, 581–591 (2015).

64. Adams, M. et al. One fly-one genome: chromosome-scale genome assembly of a single outbred *Drosophila melanogaster*. *Nucleic Acids Res.* **48**, e75 (2020).

65. Putnam, N. H. et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**, 342–350 (2016).

66. Ghurye, J., Pop, M., Koren, S., Bickhart, D. & Chin, C.-S. Scaffolding of long read assemblies using long range contact information. *BMC Genom.* **18**, 527 (2017).

67. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).

68. Dudchenko, O. et al. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000. Preprint at *bioRxiv* https://doi.org/10.1101/254797 (2018).

69. Durand, N. C. et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).

70. Denbo, S. et al. Revision of the *Capsaspora* genome using read mating information adjusts the view on premetazoan genome. *Dev. Growth Differ.* **61**, 34–42 (2019).

71. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).

72. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).

73. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158 (2020).

74. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).

75. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://arxiv.org/abs/1303.3997 (2013).

76. Open2C et al. Pairtools: from sequencing data to chromosome contacts. Preprint at *bioRxiv* https://doi.org/10.1101/2023.02.13.528389 (2023).

77. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).

78. Xu, M. et al. TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *Gigascience* **9**, giaa094 (2020).

79. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinform.* **19**, 460 (2018).

80. Walker, B. J. et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963 (2014).

81. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).

82. Laetsch, D. R. & Blaxter, M. L. BlobTools: interrogation of genome assemblies. *F1000Res.* **6**, 1287 (2017).

83. Cabanettes, F. & Klopp, C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* **6**, e4958 (2018).

84. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

85. Gertz, E. M., Yu, Y.-K., Agarwala, R., Schäffer, A. A. & Altschul, S. F. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.* **4**, 41 (2006).

86. Santini, S. et al. The compact genome of the sponge *Oopsacas minuta* (Hexactinellida) is lacking key metazoan core genes. Preprint at *bioRxiv* https://doi.org/10.1101/2022.07.26.501511 (2022).

87. Whelan, N. V., Kocot, K. M., Moroz, L. L. & Halanych, K. M. Error, signal, and the placement of Ctenophora sister to all other animals. *Proc. Natl Acad. Sci. USA* **112**, 5773–5778 (2015).

88. Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014 (2023).

89. Hoff, K. J., Lomsadze, A., Borodovsky, M. & Stanke, M. Whole-genome annotation with BRAKER. *Methods Mol. Biol.* **1962**, 65–95 (2019).

90. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

91. Brůna, T., Lomsadze, A. & Borodovsky, M. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom. Bioinform.* **2**, lqaa026 (2020).

92. Francis, W. R., Shaner, N. C., Christianson, L. M., Powers, M. L. & Haddock, S. H. D. Occurrence of isopenicillin-*N*-synthase homologs in bioluminescent ctenophores and implications for coelenterazine biosynthesis. *PLoS ONE* **10**, e0128742 (2015).

93. Townsend, J. P. et al. A mesopelagic ctenophore representing a new family, with notes on family-level taxonomy in Ctenophora:

Vampyroctena delmarvensis gen. nov. sp. nov. (Vampyroctenidae, fam. nov.). *Mar. Biodivers.* **50**, 34 (2020).

94. Babonis, L. S. et al. Integrating embryonic development and evolutionary history to characterize tentacle-specific cell types in a ctenophore. *Mol. Biol. Evol.* **35**, 2940–2956 (2018).

95. Grabherr, M. G. et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).

96. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212 (2015).

97. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

98. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).

99. The Darwin Tree of Life Project. Sequence locally, think globally: The Darwin Tree of Life Project. *Proc. Natl Acad. Sci. USA* **119**, e2115642118 (2022).

100. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

101. Potter, S. C. et al. HMMER web server: 2018 update. *Nucleic Acids Res.* **46**, W200–W204 (2018).

102. Srivastava, M. et al. The *Trichoplax* genome and the nature of placozoans. *Nature* **454**, 955–960 (2008).

103. Zimmermann, B. et al. Sea anemone genomes reveal ancestral metazoan chromosomal macrosynteny. Preprint at *bioRxiv* https://doi.org/10.1101/2020.10.30.359448 (2020).

104. Mi, H., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. Large-scale gene function analysis with the PANTHER classification system. *Nat. Protoc.* **8**, 1551–1566 (2013).

105. Lander, E. S. et al. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

106. Höhna, S. et al. RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Syst. Biol.* **65**, 726–736 (2016).

107. Köster, J. & Rahmann, S. Snakemake—a scalable bioinformatics workflow engine. *Bioinformatics* **28**, 2520–2522 (2012).

Acknowledgements

We acknowledge the support of the David and Lucile Packard Foundation, the Monterey Bay Aquarium Research Institute and the United States National Science Foundation (NSF). D.T.S. was supported by the NSF GRFP DGE 1339067; S.H.D.H. by NSF DEB-1542679; D.T.S. and O.S. by the European Research Council's Horizon 2020: European Union Research and Innovation Programme, grant no. 945026; D.S.R. by internal funds of the OIST Molecular Genetics Unit, the Chan Zuckerberg Biohub and the Marthella Foskett Brown Chair in Biology. We thank W. Patry and the staff at Monterey Bay Aquarium for providing the *Bolinopsis* specimen; and C. Dunn, J. Eizenga, R. Revilla-i-Domingo and G. Genikhovich for discussions. The *Petrosia* and *Chondrosia* genomes from the Aquatic Symbiosis Genomics project of the Tree of Life Programme, Wellcome Sanger Institute, were funded by the Gordon and Betty Moore Foundation and the Wellcome Trust.

Author information

Authors and Affiliations

1. Department of Neuroscience and Developmental Biology, University of Vienna, Vienna, Austria

   Darrin T. Schultz & Oleg Simakov

2. Monterey Bay Aquarium Research Institute, Moss Landing, CA, USA

   Darrin T. Schultz & Steven H. D. Haddock

3. Department of Biomolecular Engineering and Bioinformatics, University of California, Santa Cruz, CA, USA

   Darrin T. Schultz & Richard E. Green

4. Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA

   Steven H. D. Haddock

5. Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA

   Jessen V. Bredeson & Daniel S. Rokhsar

6. Molecular Genetics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Japan

   Daniel S. Rokhsar

7. Chan Zuckerberg Biohub, San Francisco, CA, USA

   Daniel S. Rokhsar

Contributions

D.T.S., O.S. and D.S.R. designed the scientific objectives of the study. D.T.S. planned and carried out experiments and analyses, wrote the code for the analyses and the odp package, and wrote the first draft of the manuscript. O.S. and D.T.S. performed the Bayesian analyses. D.T.S. and S.H.D.H. collected samples and created figures and tables. J.V.B. assisted in genome assemblies and identified putative centromeres. D.T.S. and D.S.R. wrote the first drafts of the manuscript with input from O.S. and S.H.D.H. All of the authors contributed to the interpretation, presentation and writing of the Article and the Supplementary Information.

Corresponding authors

Correspondence to Darrin T. Schultz, Oleg Simakov or Daniel S. Rokhsar.

Ethics declarations

Competing interests

R.E.G. and D.S.R. are paid consultants and equity holders of Dovetail Genomics. D.T.S. is a shareholder of Pacific Biosciences of California. The other authors declare no competing interests.

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Genomes of one ctenophore and two sponges.

a. The k-mer spectrum of the *Bolinopsis* data suggests that the animal is diploid, and the 1n genome size is approximately 254 Mbp. **b**. The *Bolinopsis microptera* genome assembly contains 13 chromosome-scale scaffolds, which account for 97.23% of the total bases in the assembly. Panel shows the Hi-C contact map. **c**. The cladorhizid sponge individual used in the genome sequencing at its collection site. **d**. This sponge was bioluminescent when mechanically disturbed. **e**. Its mitochondrial sequence is 99.2% identical to the previously identified bioluminescent cladorhizid sponge[36]. **f**. The estimated genome size of this sponge is 1.11 Gb, and the spectrum is consistent with diploid organisms. **g**.,**h**. Each haplotype's genome assembly has 18 chromosome-scale scaffolds based on chromatin confirmation data as shown. In haplotype A 94.2% of bases are in the chromosome-scale scaffolds. **i**. A whole-genome alignment of haplotypes A and B showed a high degree of concordance. **j**. The hexactinellid sponge collected and sequenced for this study. **k**. The estimated genome size is 1n = 141 Mb. **l**. Haplotype A contains only one haplotype of chromosome-scale scaffolds orthologous with the scaffolds of the closely-related sponge *Oopsacas minuta*. Panel shows chromatin conformation capture contact map of haplotype A. **m**. In addition to the alternate haplotype of chromosome-scale scaffolds from haplotype A, the haplotype B assembly contains the large, gene-poor, unplaced scaffolds that lack detectable homology to other sponges. The Hi-C contact map for haplotype B shown. **n**. Whole-genome alignments of the two haplotypes show colinearity. Photograph credits: (**d**.) Darrin Schultz, (**c., j.**) © 2021 MBARI.

Extended Data Fig. 2 Chromosomes are largely conserved among metazoans.

The chromosome position of orthologous proteins plotted in panels **a-n** are coloured by orthologs in the previously identified ancestral bilaterian, cnidarian, and sponge linkage groups (BCnS-ALGs)[12]. Significant inter-species chromosome pairs ($p \leq 0.05$, Bonferroni-corrected one-sided Fisher's exact test[57]) are opaque. **a**. The karyotype of the Pleurobrachiid

and lobate ctenophores is conserved ($1n = 13$). **b.-e**. Ctenophore chromosomes share macrosynteny with BCnS-ALGs, but many BCnS ALGs are split onto several ctenophore chromosomes (red dotted boxes). There are many ctenophore-specific chromosome fusions. **f.-h**. Macrosynteny is highly conserved between distantly-related demosponges. The sponge lineages shown diverged an estimated 358 Mya - 500 Mya[38]. **f.-k**. Macrosynteny is also conserved between sponge, bilaterian, and cnidarian genomes. Many chromosomes in a species of one clade have a one-to-one homologous chromosome in the other clade. The genomes of species in these clades can be described by 29 constituent BCnS-ALGs[12].

Extended Data Fig. 3 Sponge macrosynteny.

a.-c. There have been many genome rearrangements since the divergence of the demosponge *Ephydatia* and the tulip hexactinellid genome, and they share macrosynteny of only some BCnS linkage groups ($p \leq 0.05$, Bonferroni-corrected one-sided Fisher's exact test[57], opaque dots in **a**., rows in **b**., interspecies lines in **c**.). **d**. The sponge cladogram is based on Schuster et al. 2018[38]. **e**. The orthologs in A1a_x and A1a_y are predominantly present on separate chromosomes in both the tulip hexactinellid and in *Oopsacas minuta*. **f**. A1a_x and A1a_y are on partly overlapping regions of single demosponge chromosomes, but are mixed on a single *Chondrosia* chromosome. However, the linkage groups A1a_x and A1a_y are on separate chromosomes in the ctenophores and the unicellular outgroup species. This evidence suggests that hexactinellid sponges retain the ancestral state of A1a_x and A1a_y being present on separate chromosomes. The possible evolutionary scenarios explaining this karyotype will require further chromosome-scale sequencing of sponge genomes.

Extended Data Fig. 4 Seven basic ALG configurations in species quartets.

a-g. The seven configurations of ALGs found in four species highlight the evolutionary history of chromosomes. The cartoon ribbon plot in each panel shows chromosomes (horizontal bars) and the positions of genes in two

ALGs along those chromosomes (vertical blue or red lines, respectively). The cartoon Oxford dot plot in each panel shows the same information as the ribbon plot, but only in the context of the outgroup genome. The most parsimonious tree topology based on the ALG evolutionary history is also pictured.

Extended Data Fig. 5 Unicellular species chromosome-scale genome assemblies.

Hi-C heatmaps of **a**. *Salpingoeca rosetta*, **b**. *Capsaspora owczarzaki*, and **c**. *Creolimax fragrantissima* show that the assemblies are consistent with chromosome-scale assemblies of other unicellular species. **d.-e**. Genome-wide ICE-normalized [108] observed count contact maps for (**d**.) *Salpingoeca*, (**e**.) *Capsaspora*, and (**f**.) *Creolimax* are shown at MapQ0 and 10 kb resolution. Chromosome boundaries are drawn as solid black lines. The intersections of horizontal and vertical red lines mark the Centurion-estimated centromere positions. The Hi-C heatmaps of *Capsaspora* and *Creolimax* both contain inter-chromosomal hotspots that are consistent with centromeres in other species. **g.-i**. Protein orthology plots (Oxford dot plots) of the chromosome-scale genome assemblies compared to the previously published assemblies. Despite the lack of Hi-C data, the original scaffold assemblies for all three species were nearly chromosome-scale.

Extended Data Fig. 6 Visual representation of multi-species gene linkage conservation score.

a.-d. The dot plots of the *C. owczarzaki* genome show that there is conservation of the ALG_A1a linkage group in ctenophore, sponge, cnidarian, and bilaterian genomes. The conservation score can be calculated from shared gene linkages across many species. **f**. Due to the highly rearranged state of both the *Hormiphora* and *Capsaspora* genomes, a Bonferroni-corrected one-sided Fisher's exact test[57] only distinguishes three chromosome relationships as significant ($p \leq 0.05$). **e**. Calculating the orthology conservation score for the relationships in these two genomes reveals more gene linkages that have been conserved across Filozoans. Red dots here are orthologs that are in significantly-conserved ortholog

networks ($\alpha \leq 0.05$, permutation test). See complete results in Supplementary Information 11.

Extended Data Fig. 7 Filozoan and choanoflagellate genomes share macrosynteny with metazoans.

Two-way reciprocal best hits blast searches between the filasterean amoeba *Capsaspora* and animals (**a.-d**.), or between the choanoflagellate *Salpingoeca* and animals (**d.-g**.) show that the chromosomes of these unicellular species are rearranged relative to animal chromosomes, that some regions of synteny remain, and that some ALGs are split across multiple chromosomes of the unicellular species. Orthologs are coloured based on BCnS-ALGs from Simakov et al. 2022[12], and chromosome pairs with significantly-conserved macrosynteny ($p \leq 0.05$, Bonferroni-corrected one-sided Fisher's exact test[57]) have opaque dots. Axis labels show cumulative number of orthologs. Putative centromeres are marked by dotted lines.

Extended Data Fig. 8 Mixing plots of HCA-EMU-RES reciprocal best blastp results.

This figure parallels Fig. 3 of the main text, but includes more genes by requiring orthology between metazoans without requiring orthologs in corresponding outgroups. Limiting the macrosynteny search to animals shows many genes participating in the extension of metazoan ALGs to the ctenophores. The _x and _y components of ALG_Ea and ALG_G are mixed and widely distributed across single sponge chromosomes, while the (COW/SRO)-HCA-EMU-RES results show no _x and _y overlap for ALG_Ea, and little overlap for ALG_G. We placed placozoans as the sister clade to cnidarians based on the findings of Simakov et al. 2022[12]. See also Supplementary Information 13.2.2.

Extended Data Fig. 9 OrthoFinder results are consistent with the ctenophore-sister hypothesis.

a. Each green cell shows how many orthogroups support the ctenophore-sister hypothesis from each ALG in each species quartet. The Total Gene Count column is the total number of orthogroups supporting the ctenophore-sister hypothesis for that species quartet. The bottom row shows the number of unique orthogroups in each column. There are 146 orthogroups that support ctenophore-sister. **b**. The 11 orthologs that support CLA-sister in three analyses are due to a lineage-specific fission of ALG_H that is only found in the cladorhizid genome, but not in the genome of other sponges. Tree topology based on previous studies[38,109]. **c**. The *Capsaspora*-cladorhizid chromosome pairs with the most genes from ALG_H (COW4-CLA13, COW4-CLA14) are not the chromosome pairs supporting sponge-sister (magenta circles, COW4-CLA13, COW6-CLA14). **d.-h**. The fission of ALG_H is specific to the cladorhizid sponge genome and is not found in the unicellular organism *Capsaspora* (COW), in other demosponges (EMU, CRE, PFI), in cnidarians (RES), or in bilaterians (not shown). Chromosome pairs that have significantly-conserved macrosynteny ($p \leq 0.05$, Bonferroni-corrected one-sided Fisher's exact test[57]) have opaque dots.

Extended Data Fig. 10 Results of genome shuffling simulations.

a.-d. Shuffling one of the genomes before the COW-HCA-EMU-RES comparison shows that the rearranged state of the ctenophore genome, let alone the other species in the analysis, cannot explain the signal supporting the ctenophore-sister hypothesis (vertical red lines). **e.-h**. Shuffling simulations using SRO as the outgroup independently support the ctenophore-sister hypothesis. **i**. contains a legend to interpret panels a-h.

Extended Data Table 1 Linkage groups conserved in animals and unicellular outgroups
Extended Data Table 2 Ancestral linkage groups found by four-way outgroup-animal comparisons

Supplementary information

Supplementary Information

Supplementary Methods, Supplementary Tables, Supplementary Figures, Supplementary Results and Discussion, and Supplementary Notes.

Reporting Summary

Supplementary Data 1

Sequencing library details.

Supplementary Data 2

Orthologue tables.

Supplementary Data 3

GO analysis.

Supplementary Data 4

Text files of ALG mixing results.

Supplementary Data 5

Tables of ALG mixing results.

Supplementary Data 6

Bayesian analysis files.

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Schultz, D.T., Haddock, S.H.D., Bredeson, J.V. *et al.* Ancient gene linkages support ctenophores as sister to other animals. *Nature* **618**, 110–117 (2023). https://doi.org/10.1038/s41586-023-05936-6

- Received: 30 April 2022
- Accepted: 09 March 2023
- Published: 17 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-05936-6

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

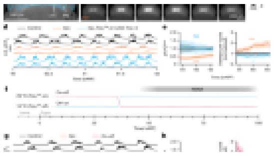
Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

Chromosomal comparisons reveal comb jellies as the sister group to all other animals

Nature Research Briefing 17 May 2023

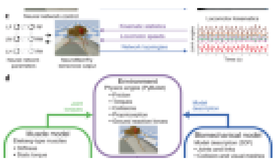
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-05936-6



A discrete neuronal population coordinates brain-wide developmental activity

09 February 2022

Bryce T. Bajar, Nguyen T. Phi, … Orkun Akin



NeuroMechFly, a neuromechanical model of adult Drosophila melanogaster

11 May 2022

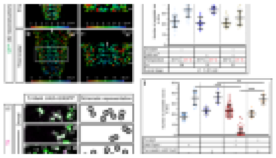
Victor Lobato-Rios, Shravan Tata Ramalingasetty, … Pavan Ramdya



Novel flight style and light wings boost flight performance of tiny beetles

19 January 2022

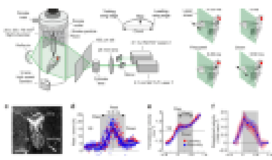
Sergey E. Farisenkov, Dmitry Kolomenskiy, … Alexey A. Polilov



[A GABAergic Maf-expressing interneuron subset regulates the speed of locomotion in Drosophila]()

22 October 2019

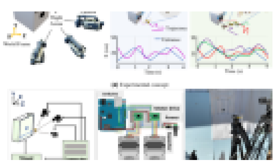
H. Babski, T. Jovanic, … A. Garcès



[Vortex trapping recaptures energy in flying fruit flies]()

26 March 2021

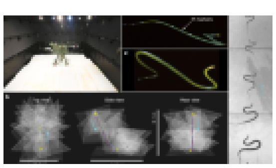
Fritz-Olaf Lehmann, Hao Wang & Thomas Engels



[Insect visuomotor delay adjustments in group flight support swarm cohesion]()

19 April 2023

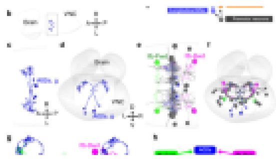
Md. Saiful Islam & Imraan A. Faruque



[Undulation enables gliding in flying snakes]

29 June 2020

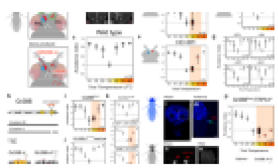
Isaac J. Yeaton, Shane D. Ross, … John J. Socha



[Regulation of forward and backward locomotion through intersegmental feedback circuits in Drosophila larvae]

14 June 2019

Hiroshi Kohsaka, Maarten F. Zwart, … Akinao Nose



[Robustness and plasticity in Drosophila heat avoidance]

06 April 2021

José Miguel Simões, Joshua I. Levy, … Marco Gallio

- Article
- Open Access
- Published: 24 May 2023

Gap junctions desynchronize a neural circuit to stabilize insect flight

- Silvan Hürkey,
- Nelson Niemeyer,
- Jan-Hendrik Schleimer,
- Stefanie Ryglewski,
- Susanne Schreiber &
- …
- Carsten Duch

Show authors

Nature **volume 618**, pages 118–125 (2023)

- 5096 Accesses
- 72 Altmetric
- Metrics details

Abstract

Insect asynchronous flight is one of the most prevalent forms of animal locomotion used by more than 600,000 species. Despite profound insights into the motor patterns[1], biomechanics[2,3] and aerodynamics underlying asynchronous flight[4,5], the architecture and function of the central-pattern-generating (CPG) neural network remain unclear. Here, on the basis of an experiment–theory approach including electrophysiology, optophysiology, *Drosophila* genetics and mathematical modelling, we identify a miniaturized circuit solution with unexpected properties. The CPG network consists of motoneurons interconnected by electrical synapses that, in contrast to doctrine, produce network activity splayed out in time instead of synchronized across neurons. Experimental and mathematical evidence support a generic mechanism for network desynchronization that relies on weak electrical synapses and specific excitability dynamics of the coupled neurons. In small networks, electrical synapses can synchronize or desynchronize network activity, depending on the neuron-intrinsic dynamics and ion channel composition. In the asynchronous flight CPG, this mechanism translates unpatterned premotor input into stereotyped neuronal

firing with fixed sequences of cell activation that ensure stable wingbeat power and, as we show, is conserved across multiple species. Our findings prove a wider functional versatility of electrical synapses in the dynamic control of neural circuits and highlight the relevance of detecting electrical synapses in connectomics.

Main

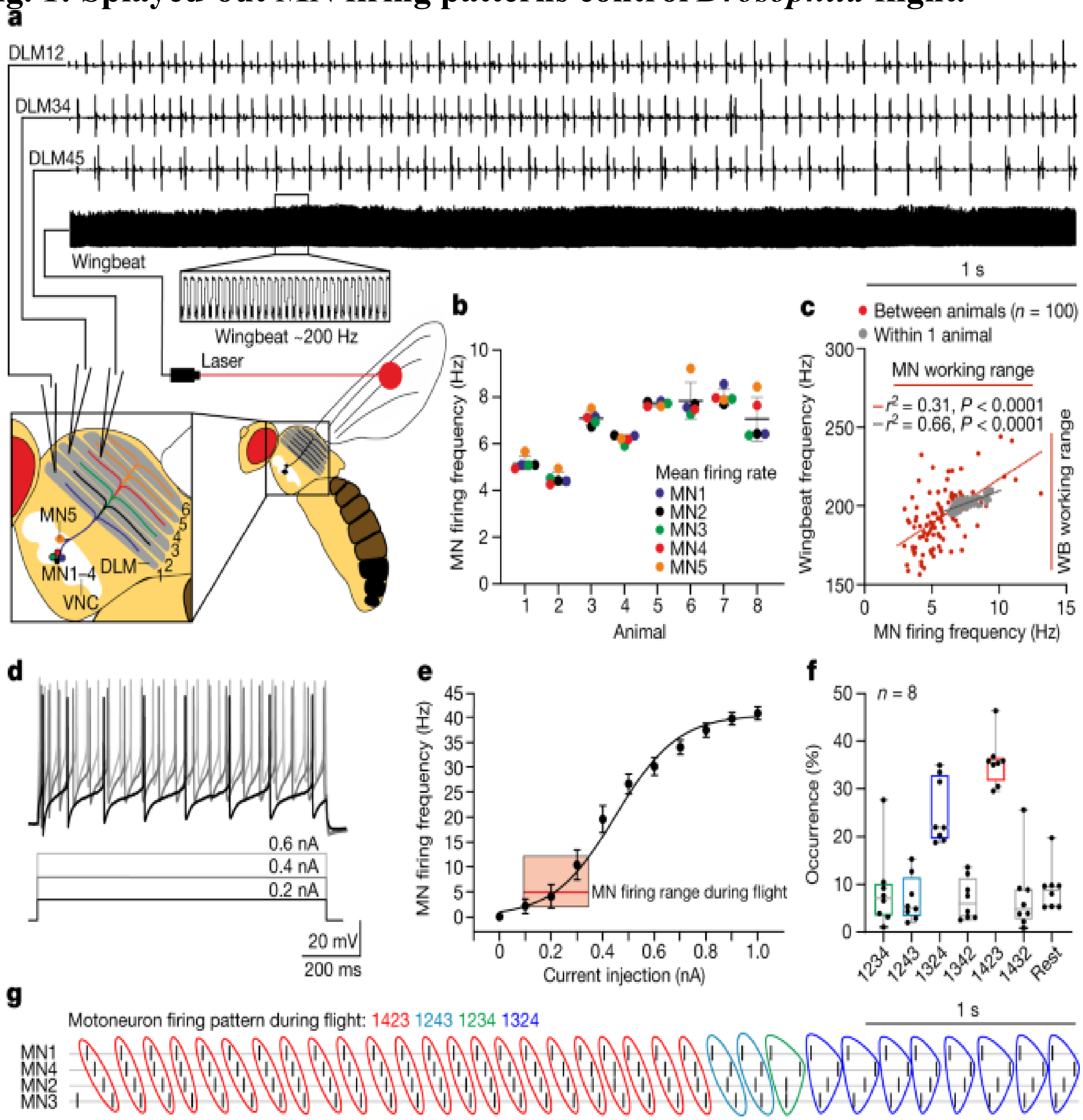
With over a million known species, insects comprise the largest group of animals on Earth[6]. Their considerable evolutionary success has been attributed to small body size and the ability to fly. These two features provide access to unutilized niches and rapid translocation, but aerodynamic constraints in small flyers require high wingbeat frequencies, and space constraints demand miniaturization of the central nervous controllers for flight[7]. In 75% of all flying insect species, highly specialized, indirect, asynchronous flight muscles form an oscillatory system that generates wingbeat frequencies of 100–1,000 Hz by reciprocal stretch activation of antagonistic wing muscles to ensure forward propulsion at low Reynolds numbers[1,8]. The flight motoneurons (MNs) that innervate asynchronous flight muscles fire at much lower frequencies, therefore not activating the muscles on a cycle-to-cycle basis. Nonetheless, power output is regulated by a CPG network in the central nervous system that controls MN firing frequencies to adjust the myoplasmic calcium levels that, in turn, regulate wingbeat frequency and amplitude[1]. Although asynchronous flight has emerged independently 7–10 times during evolution[8], neither the principles of CPG architecture for generating MN output from the miniaturized central nervous system of asynchronous flyers nor the functional consequences thereof have been identified.

Asynchronous flight motor patterns

To quantify asynchronous flight patterns and decipher CPG architecture, we used the firing output of the five identified MNs (MN1–5) innervating the dorsal longitudinal wing depressor muscle (DLM) of the genetic model system[9,10,11] *Drosophila melanogaster* as well as other insect species (to test for generality). The DLM provides the force for wing downstroke, consists

of six muscle fibres, each of which is innervated by one of five identified MNs[9,10,11] (MN1–5; Fig. 1a). MN1–4 each target one of the four most ventral DLM fibres ipsilateral to their somata, whereas MN5 innervates DLM fibres 5 and 6 on the side contralateral to the MN5 soma (Fig. 1a). This neuromuscular architecture is conserved across insect species examined (locust[12], moth[13], blowfly[14]).

Fig. 1: Splayed-out MN firing patterns control *Drosophila* flight.



a, Representative recording of MN1–5 and wingbeat frequency (bottom trace, magnified in the black box) during tethered flight. Colour-coded

schematic of MN1–5 in the VNC and axonal projections to the six fibres of the DLM. **b**, Average MN firing frequencies are similar within each animal. The colour code is the same as in **a**. $n = 8$ animals. Data are mean ± s.d. **c**, MN firing frequency and wingbeat (WB) frequency (the red bars indicate working ranges) are linearly related within an animal (grey dots; correlation coefficient, $r^2 = 0.63$; $P < 0.0001$, two-tailed t-test) and with larger variance also across animals (red dots; $n = 100$; correlation coefficient, $r^2 = 0.31$; $P < 0.0001$, two-tailed t-test). **d**, The firing responses of MNs (top traces) to current injections of different amplitudes (bottom traces). **e**, The mean MN response frequency (f) and injected current amplitude (I) are approximately linearly related for 2–30 Hz ($n = 15$ animals), therefore exceeding the normal MN firing frequencies observed during flight (inset; ~3–12 Hz, data from **c**). Data are mean ± s.e.m. (main plot) and median ± range (inset). **f**, During flight, MN1–4 spikes are dispersed in time (in splay state) with characteristic sequences. Each animal switches between different splay states during flight, but the same splay states are preferred across individuals ($n = 8$). The box plots show the median (centre line), quartiles (box limits) and range (error bars). **g**, Timing of MN1–4 spikes in four subsequent splay states (1423 (red); 1243 (turquoise); 1234 (green); 1324 (dark blue)) during flight.

In vivo recordings of MN1–5 from their DLM target muscle fibres using simultaneous laser-based wingbeat detection during tethered flight show that each MN fires only every approximately 20th to 40th wingbeat (Fig. 1a–c). Although the firing frequencies of MN1–5 can vary between animals and are adjusted on demand[1,15,16], within a given animal and power demand, all five MNs always fire at the same frequencies (Fig. 1b) and with similar variance in the interspike interval (ISI; coefficient of variation is 0.22 ± 0.07 (mean ± s.d.) for each MN).

MN1–5 firing frequencies directly control muscular tension and stretch activatability by adjusting myoplasmic calcium levels and, therefore, wingbeat frequency and stroke amplitude[1,17]. Accordingly, in single animals, alterations in power demands go along with changes in MN1–5 firing frequency that are proportional to wingbeat frequency changes (Fig. 1c (grey dots)). A linear relationship between average MN firing and average wingbeat frequency within the normal flight working range was further confirmed by recordings in 100 animals, although interanimal variability

yields a lower correlation between MN firing rate and wingbeat frequency (Fig. 1c (red dots)). This correlation is increased when analysing the changes in the MN firing rate in relation to changes in the wingbeat frequency (Extended Data Fig. 4e–g). Thus, the central nervous system controls asynchronous flight muscle power output neither by the recruitment of different motor units, nor on the scale of single wingbeats, but the frequency of MN1–5 population firing is the key regulator of wing power production[1]. We show that MN excitability is well suited to dynamically regulate wing power output. First, MN1–5 respond to constant input with slow tonic firing (Fig. 1d). Second, a nearly linear input–output relationship as observed in *f*–*I* curves (3–30 Hz; Fig. 1e) covers and even exceeds the working range observed during tethered flight (approximately 3–12 Hz; Fig. 1c). The excitability of MN1–5 is therefore tuned to linearly translate synaptic input into tonic firing to regulate wingbeat in the working range of flight.

Despite equal firing frequencies at any given wingbeat frequency[1,15] (Fig. 1a,b), the five MNs in this system do not fire in synchrony. Instead, as previously suggested[18], their spikes are splayed-out in time with firing phases across neurons dispersed approximately equidistantly, resulting in stereotyped preferred sequences of MN1–5 spiking (Fig. 1g), which we name splay states. A network splay state means that each individual cell in a network of *N* neurons fires regularly, yet with a constant, non-zero phase relationship with respect to the firing of each other neuron. Firing phase differences between neurons approximately correspond to 1/*N*th of a neuron's ISI or a multiple ($<N$) thereof. Importantly, the sequence of MN1–5 firing can change intermittently, but the CPG robustly slides back into one or two of the most preferred splay states. Notably, the same splay states are preferred across animals (Fig. 1f), indicating that they result from hard-wired CPG circuitry. Splayed-out firing results in characteristic phase relationships between each pair of MN1–5. These phase relationships are conserved across individuals (Extended Data Fig. 1a–c). The phase relationships of all MN1–5 pairs show a gap around phase zero, indicating out-of-phase firing. Given that multiple splay states exist during flight (Fig. 1f,g), firing of MN pairs is not precisely phase locked (Extended Data Fig. 1a–c). Yet, restricting phase histograms to one splay state narrows the phase relations (Extended Data Fig. 1d). Moreover, synchronous spikes between

MN pairs occasionally occur during switches between different splay states, but not within a stable splay state (Extended Data Fig. 1d).

Our characterization of phase relationships between pairwise combinations of the MN1–5 enables us to test for across-species conservation of the motor patterns (see Supplementary Videos 1–3 for the flight of three species). This reveals a notable similarity between *D. melanogaster* and the gold fly *Lucilia* sp. (Extended Data Fig. 2a,b). Similar phase relationships have also been suggested for *Calliphora erythrocephala*[14], *Eucaliphora lilaea* and *Musca domestica*. Moreover, using the MN4–MN5 pair (Extended Data Fig. 2c), our measurements indicate conservation of CPG architecture between two *Drosophila* species (*D. melanogaster* and *Drosophila hydei*), other dipteran genera (*Calliphora* and *Musca*) and, to a certain degree, even between *Diptera* (for example, flies) and *Hymenoptera* (for example, honey bee). Given that asynchronous flight evolved multiple times independently[8], splayed-out firing has probably provided selective benefits over millions of years. To gain mechanistic and functional insights into the CPG network that controls asynchronous flight, we next analysed the network principles underlying splay state motor pattern generation and the resulting functional benefits for flight performance.

Network splay states produced by a minimal CPG

From invertebrates[19] to mammals[20], the timing and pattern of MN activation during locomotion typically rely on networks of premotor interneurons. By contrast, we provide evidence that splayed-out MN1–5 firing is generated by interactions between the MNs themselves. First, unpatterned optogenetic activation of excitatory, cholinergic input to MN1–5 increases MN firing frequencies without changing their phase relationships, preserving a characteristic gap around phase 0 (Extended Data Fig. 3). We confirmed both physiologically[21] and anatomically that MNs receive excitatory cholinergic input to their dendrites (Extended Data Fig. 4a,b), and that all MNs share common synaptic input (Extended Data Fig. 4c,d) as previously proposed[18]. Tonic common excitatory synaptic drive to MN1–5, as induced by optogenetic stimulation, is therefore transformed into out-of-phase MN1–5 firing. Second, unpatterned, tonic, optogenetic stimulation of the five MNs (see Extended Data Fig. 5a,b for selective expression of transgenes in MN1–

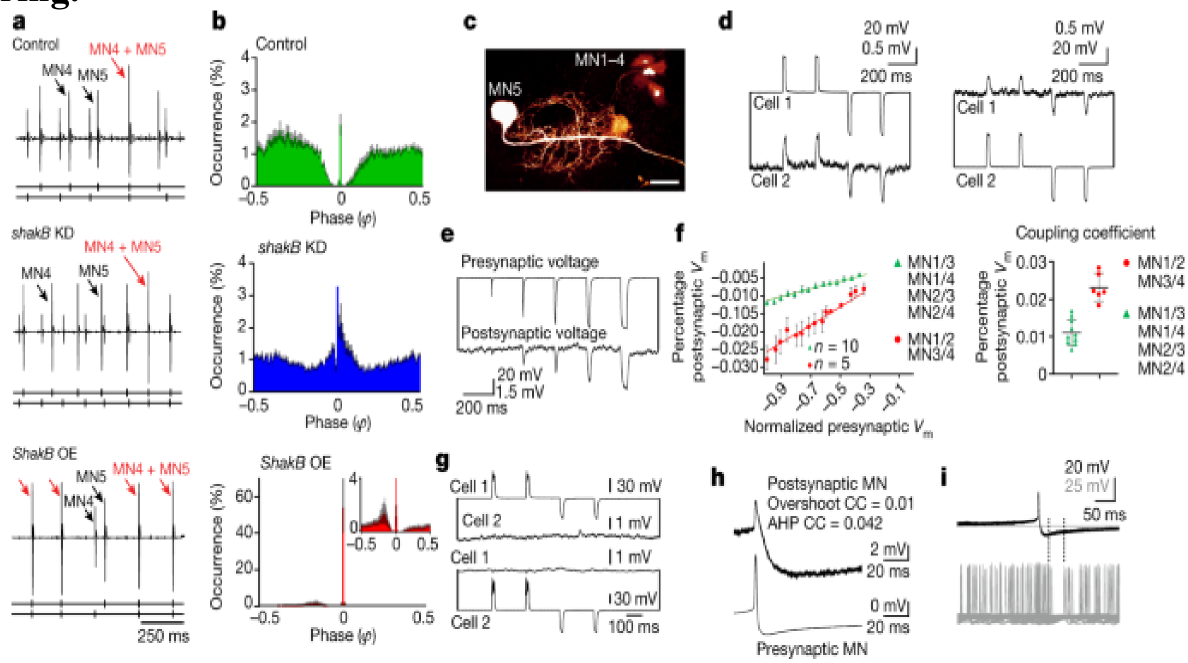
5) during flight increases their firing frequency and, therefore, wingbeat power, but it does not change phase relationships between MNs, as exemplified by the same characteristic gap around phase 0 for the MN4–MN5 pair (Extended Data Fig. 5). The strength of this dip around phase 0 in MN phase histograms is strongly correlated with full network splayness (definitions are provided in the Methods) in both simulated and experimental data (Extended Data Fig. 6). Similar phase histograms of MN pairs before and during optogenetic stimulation therefore indicate unaltered splayness. The generation of continuous splayed-out firing after selective, unpatterned, optogenetic stimulation of all five MNs indicates that interactions between MNs shape the firing patterns. An alternative possibility would be synaptic output from MN1–5 to feedback interneurons, which then generate the pattern. However, neither trans-synaptic mapping nor genetic markers for central output synapses from MN1–5 provide evidence for postsynaptic partners in the ventral nerve cord (VNC; Extended Data Fig. 7). Thus, it seems probable that the timing and pattern of MN activation does not require patterned activity of interneurons and, instead, the MN1–5 ensemble constitutes a minimal CPG.

One possible mechanism to transform common excitatory input to an ensemble of MNs into dispersed firing is lateral inhibition among MN1–5 by chemical synapses[15]. We reject this possibility by combining targeted genetic manipulation of MNs with in vivo recordings during flight. Knockdown (KD) of receptors for inhibitory transmitter (GABA A receptors (GABA-ARs) and glutamate-gated chloride channel (GluCl)) increases MN firing frequencies, but the phase relationships remain similar (Extended Data Fig. 6e,f). Lateral inhibition through chemical synapses is therefore not required for pattern generation.

Another possibility to create a neural network exclusively from MN1–5 is to connect them with electrical synapses. This has previously been suggested[22,23,24], but experimental evidence has been lacking, and electrical coupling seems difficult to reconcile with desynchronized firing. We have tested this by genetic manipulation of innexins[25], the invertebrate counterparts of connexins[26], which comprise the pore-forming proteins of electrical synapses. ShakingB (ShakB) is the innexin expressed in the *Drosophila* escape circuit, including in MN1–5. KD of *shakB* using RNA

interference (RNAi) in MN1–5 reduces electrical coupling below the detection threshold in paired patch-clamp recordings (Fig. 2g).

Fig. 2: Electrical synapses shape CPG output by desynchronizing MN firing.



a, Representative recording of MN4 and MN5 during tethered flight in the control (top) and after RNAi KD of *shakB*-encoded electrical synapses (DLM-Split-GAL4>UAS-shakB-RNAi, middle) and overexpression of ShakB in MN1–5 (bottom). The black arrows mark MN4 and MN5 spikes, and the red arrows indicate simultaneous MN4–MN5 spikes. **b**, Phase histograms of the occurrence of MN5 spikes (*y* axis) in relation to consecutive MN4 spikes (phase $\varphi = 0$ corresponds to the MN4 spike) for control (top), *shakB* KD (middle) and ShakB overexpression (bottom) with a magnified view (inset) ($n = 10$ animals for each genotype). Data are mean (coloured bars) ± s.e.m. (grey). **c**, MN1–5 dye coupling in the *dfmr1* RNAi KD background to increase dye uptake[28]. Scale bar, 20 µm. **d**–**h**, Intracellular recordings of MN pairs. **d**, Hyperpolarizations and depolarizations were conducted bidirectionally (from cell 1 to cell 2 and vice versa). **e**, Increasing current injection amplitude (top trace) increases response amplitudes in electrically coupled MNs (bottom trace). **f**, Plotting the mean presynaptic voltage (V_m) area against the mean postsynaptic voltage area (left) reveals linear relationships, but regression slopes differ

between distinct MN pairs (5 animals with strong coupling between the MN1–MN2 and MN3–MN4 pairs; 10 animals with weak coupling between the MN1–MN3, MN1–MN4, MN2–MN3 and MN2–MN4 pairs). The mean CC (postsynaptic peak voltage divided by presynaptic peak voltage; right) differs significantly between MN1–MN2, MN3–MN4 (red, 6 pairs) and all other pairs (green, 10 pairs). Data are mean ± s.e.m. (left) and mean ± s.d. (right). **g**, RNAi KD of *shakB* eliminates detectable electrical coupling between MNs (3 animals). **h**, The CC for the spike AHP is higher than for the spike overshoot. **i**, Firing of a coupled MN ceases during the AHP of the presynaptic MN.

Genetic manipulation of *shakB* in the five MNs disrupts the splay state without affecting the firing frequencies (Fig. 2a,b). As exemplified for the MN4–MN5 pair, in control animals ($n = 7$), firing of MN5 is inhibited before and after the occurrence of MN4 spikes (Figs. 2a,b (top)), therefore resulting in a characteristic gap around phase 0 (Extended Data Figs. 3b and 5e) that is indicative for network splayness (Extended Data Fig. 6a–d). By contrast, with *shakB* KD in MN1–5, the gap around phase 0 is absent and firing is more random with a slight preference for in-phase firing (Fig. 2a,b (middle)). Thus, electrical synapses are required for firing desynchronization in this small CPG. Recordings of all five MNs during flight confirm that *shakB* KD in MNs impairs their normal phase relationships and, therefore, the splay state (compare Extended Data Fig. 8 with Extended Data Fig. 1). This contradicts the common notion that electrical synapses function to synchronize network activity[27].

Key to this role of gap junctions is weak electrical coupling, because increasing electrical coupling by overexpression of ShakB in MN1–5 causes firing synchronization (Fig. 2a,b (bottom)). Thus, the in vivo data indicate that weak electrical synapses cause firing desynchronization, but stronger coupling synchronizes firing. Weak electrical coupling between MN1–5 is also supported by anatomical experiments. Although we have previously not observed diffusion of small dye tracer molecules through gap junctions between MN1–5[16,21], in *Drosophila* neurons, dye uptake during iontophoresis and subsequent diffusion through *shakB*-encoded electrical synapses are increased by KD of the fragile-X mental retardation protein

(FMRP) (encoded by *dfmr1*)[28]. In this genetic background dye coupling of MNs is reliably observed (Fig. 2c).

In summary, a minimal CPG of five electrically coupled MNs is sufficient to pattern splayed-out firing across power demands and does not rely on additional interneurons or chemical synapses. First, lateral inhibition by chemical synapses is not required (Extended Data Fig. 6e,f); second, we found no evidence for chemical output synapses from MN1–5 in the VNC (Extended Data Fig. 7); third, unpatterned optogenetic activation of either presynaptic cholinergic neurons or the MNs during tethered flight increases MN1–5 firing rates and wingbeat frequency (Extended Data Figs. 3 and 5), but the phase relationships between MNs remain unaltered (compare Fig. 2a with Extended Data Figs. 3b and 5e). Finally, weak electrical synapses between MNs are required to generate normal phase relationships (Fig. 2a,b). However, this raises the question of how electrical synapses cause firing desynchronization.

Electrical coupling within the CPG

A first step in deciphering the mechanism is to characterize the electrical synapses. Dual in situ patch-clamp recordings of MN pairs confirm electrical coupling and demonstrate that the electrical synapses are weak, bidirectional and non-rectifying (Fig. 2d). Non-rectifying, because depolarizing and hyperpolarizing current injections (20 ms duration) into one MN cause gap junctional potentials in the other MN. The relationship between presynaptic and postsynaptic charge is linear (Fig. 2e,f). Bidirectional, because the direction of charge transfer can be reversed (Fig. 2d). Compared with common CCs (postsynaptic charge/presynaptic charge[29]) estimated in different types of neurons[29] (0.02–0.2), electrical synapses between the five MNs are weak, but coupling is twice as strong for the MN1–MN2 and MN3–MN4 pairs (coupling coefficients (CC) = 0.023 ± 0.003) compared with all of the other possible combinations of MN1–4 pairs (CC = 0.01 ± 0.0027; Fig. 2f).

The possibility that electrical synapses may desynchronize neural networks under specific conditions has been suggested by theoretical work[30,31,32,33]. Experimentally, a transient desynchronization of electrically coupled

cerebellar Golgi cells has been described in one study, but for the specific condition of sparse input[34]. There, the transient desynchronization is attributed to a more effective gap junctional transmission of the afterhyperpolarization (AHP) of the action potential compared with its depolarizing overshoot[34]. Similarly, we found that the AHP of MN spikes is transmitted more effectively (CC, 0.042 ± 0.04) through ShakB-mediated electrical synapses than the brief spike overshoot (CC, 0.01 ± 0.004; Fig. 2h). CCs can differ for different components of the action potential because the duration of the presynaptic signal and the time constant of the postsynaptic membrane shape the junction potential[29]. Paired in situ current-clamp recordings of MNs that were induced to fire tonically by somatic current injection indicate that firing of one MN can depress firing of the other during and shortly after the AHP (Fig. 2i). However, the preconditions for desynchronization, as described for electrically coupled cerebellar Golgi cells[34], are probably not fulfilled in the insect asynchronous flight CPG. First, network desynchronization by electrical synapses during flight is not transient but permanent. Second, network desynchronization manifests not only under sparse synaptic input regimes, but through the full range of synaptic input that occurs during flight. Thus, although we cannot rule out a contribution of the AHP, a general mechanism for small network desynchronization without a pronounced AHP (which is probably not sufficient; see above) or inhibitory synapses (not necessary for MN desynchronization; Extended Data Fig. 6e,f) is required. Indeed, theoretical considerations of network connectivity as well as cell-intrinsic excitability provide insights into how the observed splay states are generated.

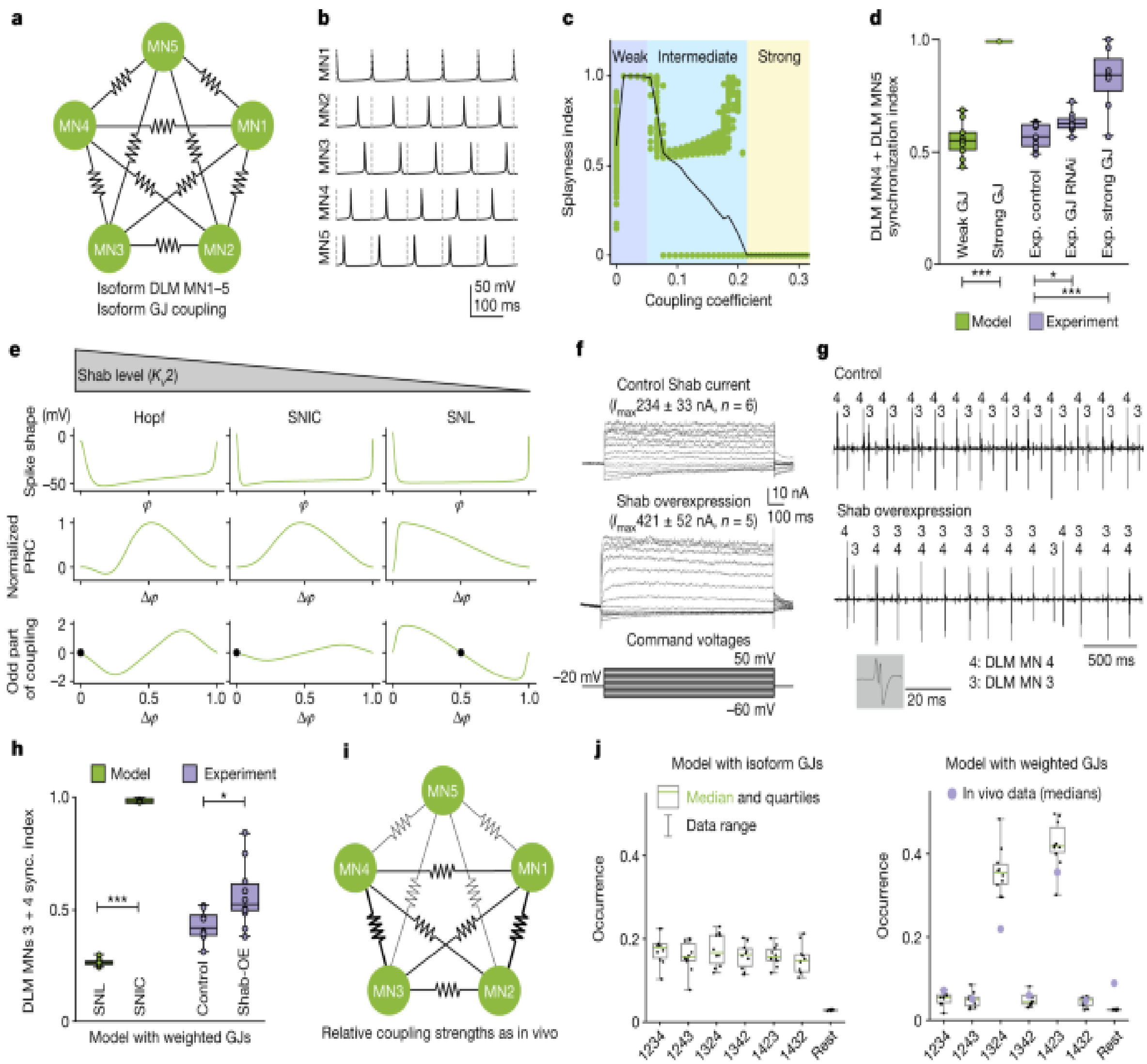
Coupling strength and excitability class

According to the theory of coupled phase oscillators, splayed-out network states are obtained if pairs of neurons have a preference to fire out of phase, that is, they are phase-repulsive[24]. However, for networks of more than two neurons, a strict antiphase state that maximizes the difference between the firing phases of two neurons cannot be achieved across all neuronal pairs; these systems are therefore called frustrated[35,36]. Typically, they can organize into a splay state that equidistantly maximizes phase distances among neurons to minimize frustration. The phase difference between

individual pairs of neurons is therefore not maximal (that is, smaller than half the ISI), but firing phases of neuron pairs are still separated across the population of cells, which reflects the splay state that we observe in vivo (Fig. 1a,f,g and Extended Data Figs. 1 and 6a–d).

Whether pairs of neurons tend to fire out of phase depends on the specific combination of synaptic connectivity and cellular excitability (in other words, the dynamics of action potential generation). Cellular excitability classes characterize qualitatively different types of spiking dynamics and are associated with distinct mathematical bifurcation types at spike onset (that is, threshold). Given the weak electrical coupling among MN1–5, in theory, repulsive phase coupling[24,33,37,38] can be fostered by specific excitability classes. To examine this hypothesis, we use a three-dimensional conductance-based neuron model fitted to MN sodium and delayed rectifier ion channel kinetics[39] (Methods) to generate a gap junctional network of MN1–5 (Fig. 3a). Subjecting the single neuron model to a mathematical bifurcation analysis within the physiological parameter range identified a cellular excitability class that our analysis predicts to favour out-of-phase firing (through a homoclinic spike-onset near the saddle-node loop (SNL) bifurcation; Methods and Extended Data Fig. 9a). Indeed, five identically, weakly electrically coupled MNs with cellular dynamics near the SNL point robustly exhibit a desynchronized splay state (Fig. 3b). Note that the minimal conductance-based model does not contain a pronounced AHP (Fig. 3b (spike shape)), showing that the presynaptic AHP is not mandatory for network desynchronization.

Fig. 3: Weak electrical synapses and saddle homoclinic orbit (HOM) excitability dynamics underlie firing desynchronization.



a,**b**, Conductance-based model with identical gap junctions (GJs, coupling coefficient CC < 0.05) between MN1–5 (**a**) produces splayed-out firing (**b**). **c**, Weak CCs (<0.05) cause high splayness indices (Methods); intermediate CCs (0.05–0.21) decrease splayness and yield multiple stable network states; CCs of >0.21 yield network synchrony (splayness = 0; 200 simulations per CC, green dots; the black line shows the average). **d**, Quantification of 60 s simulations (green, 10 simulations per condition) with a weak CC (0.005) yields significantly lower MN4–MN5 synchronization indices (Methods; median = 0.54) compared with strong coupling (CC = 0.258; median = 1.0; two-sided Mann–Whitney *U*-test, $P = 0.0002$). Similarly, experimental (exp.; purple) synchronization indices are significantly lower in the controls (median = 0.56, 11 animals) compared with the ShakB-overexpression group (median = 0.84, 7 animals, $P = 0.0008$). **e**, Increasing Shab channel levels

transforms the single-MN model dynamics from HOM near the SNL point through SNIC to Hopf types (Extended Data Fig. 9a), which vary in action potential waveform (top) and PRC (middle). Averaging theory (Methods and equation (4)) yields the odd part of the coupling function (bottom) for each phase distance $\Delta\varphi$, of which the fixpoints (black dots) determine stable network states. Only the SNL type yields one stable fixpoint at phase 0.5, therefore favouring anti-phase firing. **f**–**h**, Shab overexpression in MN1–5 nearly doubles Shab current (**f**), which causes in-phase firing of the MN3–MN4 pair (**g**) and significantly (two-sided Mann–Whitney *U*-test, $P = 0.0434$) increased synchronization indices (median = 0.52, 10 animals) compared with the controls (median = 0.43, 8 animals) (**h**). Similarly, increasing Shab in models with weighted GJs (as in Fig. 3i) significantly increased MN3–MN4 synchronization (sync.) indices (median = 0.99, $P = 0.0002$). **i**,**j**, Network models with heterogenous CCs (**i**) as found in vivo (Fig. 2f) yield preferred splay states as in animals (**j**, right; 10 simulations per condition; 8 animals, the purple dots depict median values of in vivo data), whereas homogenous coupling does not (**j**, left). For the box plots in **d**, **h** and **j**, the median (centre line), quartiles (box limits) and range (whiskers) are shown.

By contrast, systematic variation in electrical coupling strength reveals that weak electrical coupling is required because models with non-zero, small CCs (<0.05), as obtained in vivo (Fig. 2f), yield splayed-out firing (Fig. 3c), whereas, for CCs between 0.05 and 0.21, network in-phase synchronization increases and, for CCs of >0.21, the network state is synchronized (Fig. 3c). To test these model predictions experimentally, we manipulated gap-junction strength genetically and quantified the synchrony of the MN4–MN5 pair from in vivo recordings during flight (Fig. 3d). In control animals with weak gap junctions, the synchronization index (Methods) is low, similar to model simulations with weak gap junctions. RNAi KD of electrical synapses increases synchronization in vivo (Fig. 3d), underscoring that gap junctions are required for desynchronization. Finally, the model predicts synchrony for strong electrical coupling. Indeed, strengthening of electrical coupling by ShakB overexpression significantly increases synchrony in vivo (Fig. 3d; see also Fig. 2b). Thus, weak electrical coupling is required for network desynchronization.

The mechanistic core of the splay state

We next examined why cellular excitability dynamics close to the SNL point favour splay states. For regularly firing cells with all-or-none action potentials, three main dynamical excitability classes—shaped by cell-intrinsic properties including the ion channel composition—exist[40]. Mathematically, these correspond to three distinct spike-onset bifurcations: the subcritical Hopf, the saddle-node-on-an-invariant cycle (SNIC) and the saddle homoclinic orbit (HOM) bifurcations (Fig. 3e (top row)), which in turn qualitatively determine the phase dependence of a neuron's sensitivity to inputs (also termed the phase-response curve (PRC)) in a bifurcation-specific and, therefore, excitability-class-specific manner (Fig. 3e (middle row)). For the network state, the coupling function, which combines the influence of the PRC and electrical coupling, is decisive (definitions are provided in the Methods), therefore assigning a direct impact on network dynamics to both gap junctions and cellular voltage dynamics[38]. Specifically, an asymmetry of the PRC shape favours stable phase relationships of spiking in the network; these are predicted by the fixpoints of the coupling function's odd part (that is, fixpoints that can be read off the asymmetric components of the function; Methods). Our analysis shows that, for weak reciprocal electrical coupling, out-of-phase firing of identical neurons can be obtained only in combination with a PRC that is monotonically decreasing around phase 0.5 (a graphical explanation is provided in Extended Data Fig. 9b,c). Such a PRC is found for HOM excitability, including dynamics close to the border of SNIC firing, the latter corresponding to the SNL point introduced above (Fig. 3e (middle row)). Here, weak electrical coupling combined with such cellular dynamics results in a stable fixpoint at out-of-phase firing (Fig. 3e (bottom row)). By contrast, for the other two spike-onset bifurcations, synchronous in-phase firing is expected (compare also the stable fixpoints for the phases between neurons; Fig. 3e (bottom row)). Precisely, we predict that MN1–5 should dynamically reside close to the SNL point. Here, the PRC shape enables a network splay state, yet the slope of the neurons' firing-rate versus current-input (*f*–*I*) curves is not as steep as deeper in the HOM regime, therefore enabling smooth control of MN firing frequency and wingbeat power, as observed in vivo (Fig. 1c–e).

To test this model prediction, the excitability class of MN1–5 has to be manipulated in vivo. Shab potassium channels constitute around 50% of the delayed rectifier current in these MNs (Extended Data Fig. 10). Bifurcation analysis of the single-cell MN model reveals an SNL bifurcation (Extended Data Fig. 9a). Moreover, increases in total Shab channel conductance, g_{shab}, induce several transitions in excitability class[37], from HOM (at low g_{shab}), through the SNL point into a SNIC (medium g_{shab}) and a Hopf bifurcation (high g_{shab}, Fig. 3e and Extended Data Fig. 9a). At both excitability dynamics resulting from higher g_{shab} (SNIC and Hopf), the weakly coupled MN1–5 network exhibits a synchronized state, whereas SNL dynamics favour splay states (as described above). We tested this model prediction experimentally by genetic manipulation in vivo. Targeted overexpression of Shab in MN1–5 causes a near doubling of Shab delayed rectifier current (Fig. 3f) without compensatory changes in the amplitudes of non-Shab delayed rectifier current (Extended Data Fig. 10). Recordings of the MN3–MN4 pair were used to test whether this caused firing synchronization in vivo. In controls, both MNs fire out of phase (Fig. 3g (top) and Extended Data Fig. 1). By contrast, with Shab overexpression, the MN3–MN4 pair shows markedly increased synchronization (Fig. 3g (bottom)). Variability in synchronization strength between animals (Fig. 3h) is probably caused by different overexpression strengths and variable levels of other delayed rectifier channels. We also cannot exclude some compensatory regulation of yet other currents. However, pooling the data from all recordings ($n = 10$) revealed a significant increase in firing synchrony after overexpression of Shab (Fig. 3h), as predicted by an increase in g_{shab} in the network model (Fig. 3h). By contrast, our model predicts that a decrease in g_{shab} shifts MNs deeper into the HOM regime but does not cause a transition in excitability class (Fig. 3e and Extended Data Fig. 9a). Thus, with reduced g_{shab} in MN1–5, the network model predicts desynchronization (Fig. 3e). We tested this prediction experimentally. Compared with the control ($n = 7$ animals) targeted RNAi KD of *Shab* in MN1–5 ($n = 11$ animals) reduced the Shab current by around 70% (Extended Data Fig. 10). Paired in vivo recordings of MN4 and MN5 during flight confirm a low synchronization index as predicted by the model (median synchronization indices are 0.559 for control flies ($n = 11$ animals) and 0.598 for *Shab* KD ($n = 11$) and were not statistically different; $P = 0.393$, $U = 74$, two sided Mann–Whitney U-test).

The theoretical principles underlying this mechanism are independent of model details (instead they depend on the dynamical excitability class). Weak electrical coupling between neurons with SNL excitability therefore provides a generic mechanism that suffices to produce the observed desynchronized (splay) states in small networks. Importantly, this mechanism holds across different firing rates and does not require a pronounced AHP, but the AHP may further stabilize desynchronization. Although previous studies have provided mechanisms by which networks with electrically coupled neurons can be configured into desynchronized, phase-locked firing[41]—such as inhibitory synapses overriding the synchronization by electrical synapses[42] or interacting with the spiking synchrony[43,44], silencing of electrical synapses in a network with mixed chemical and electrical synapses[42,45], or strongly asymmetric electrical synapses and specific input regimes[46]—desynchronization by the electrical synapse alone, without the need of additional chemical synapses, other network motives or feedback from the network, is conceptually new.

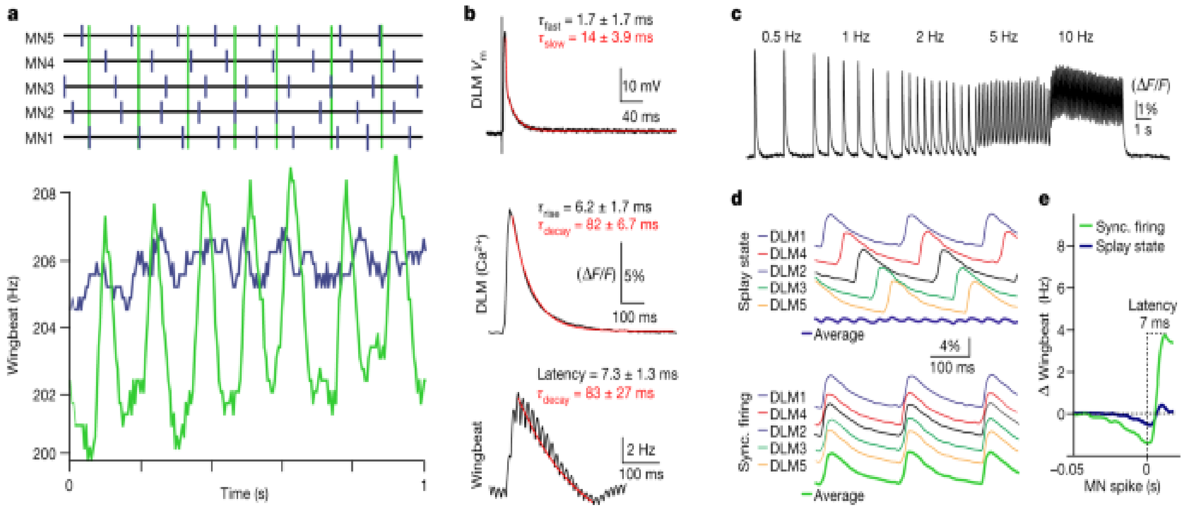
Sequence preference requires heterogenous coupling

Our in vivo recordings show that MN1–5 firing is not only desynchronized, but organized into preferred sequences of cellular activation (within the splay state; Fig. 1f,g). This raises of the question of how the preferred sequences are generated. Our modelling (5 min simulations in network models with SNL neurons and noise) demonstrates that homogenous electrical coupling does not show sequence preference (Fig. 3j (left)). By contrast, heterogenous electrical coupling between MN1–5 (Fig. 3i) as observed in situ (Fig. 2f) produces the same preferred splay states with similar sequence statistics (Fig. 3j (right)) as observed in vivo across animals (Fig. 1f). We conclude that heterogenous weak electrical coupling among neurons with HOM excitability (near SNL) sufficiently explains the preferred splay states in this small network.

Splay state serves stable wing power

MN1–5 splayed-out firing in multiple species suggests that it provides a functional benefit. We tested this hypothesis by comparing the wingbeat frequency between normal splay state and synchronous MN firing, the latter induced by increasing electrical coupling through genetic manipulation (Figs. 2a and 3c). In vivo, synchronous MN firing causes 8× higher wingbeat frequency fluctuations over time compared with firing in splay state (Fig. 4a). The amplitudes of such fluctuations are close to changes in wingbeat frequency in response to optomotor input[1,16] and are therefore functionally relevant for flight altitude control. In the steady-state, muscular stretch sensitivity and wingbeat frequency are directly proportional to the myoplasmic calcium levels[1], but the dynamics of myoplasmic calcium in DLM fibres is unclear. To link MN firing patterns to wingbeat frequency, we measured the kinetics of muscle fibre membrane potential and myoplasmic calcium concentration changes after MN1–5 firing (Fig. 4b). An MN spike causes a rapid depolarization of the DLM fibre membrane (Fig. 4b (top)), which is followed by a myoplasmic calcium signal with a rise time constant of around 6 ms and a decay time constant of about 80 ms (Fig. 4b (middle)). If muscular stretch sensitivity was controlled by the myoplasmic calcium levels with a ms temporal resolution, changes in wingbeat frequency during simultaneous MN firing should follow the same time course as changes in myoplasmic calcium across all six DLM fibres. This is precisely what we found (Fig. 4b (bottom)). Synchronous MN firing causes peak wingbeat frequencies after 7.3 ± 1.3 ms (myoplasmic calcium rise time constant τ = 6.2 ± 1.7 ms) that decline with a time constant τ = 83 ± 27 ms (calcium decay time constant is 82 ± 6.7 ms).

Fig. 4: Splayed-out firing ensures stable wingbeat power.



a, Representative traces of wingbeat and MN multiple-unit recordings during flight (7 animals in each condition). Synchronous MN firing (top, green) causes around eight times larger fluctuations in wingbeat frequency (bottom, green) compared with splayed-out firing (blue). **b**, DLM fibre voltage response to one MN spike is fast (top; decay τ_{fast} ~ 1.7 ms, τ_{slow} ~14 ms; 3 animals). Both the resulting myoplasmic calcium signal (middle; GCaMP8f $\Delta F/F$; rise $\tau = 6.2 \pm 1.7$ ms; decay $\tau = 82 \pm 6.7$ ms; 7 animals) and wingbeat frequency changes after MN spiking (bottom; latency = 7.3 ± 1.3 ms; decay $\tau = 83 \pm 27$ ms; 7 animals) follow similar slower time courses. **c**, DLM6 calcium responses to different MN5 firing frequencies (Supplementary Video 4). **d**, As each DLM fibre is innervated by one MN, splayed-out firing must result in time-shifted calcium responses (top), but synchronous MN firing results in time-locked calcium responses across fibres (bottom). The resulting average myoplasmic calcium levels across all fibres is considerably larger for synchronous (bottom panel, bottom trace) compared with for splayed-out firing (top panel, bottom trace). **e**, As a result, during flight, splayed-out firing is accompanied by small transient fluctuations in wingbeat frequency (blue; $n = 7$ animals) but synchronous firing is accompanied by significantly ($P = 0.0010$, two-sided unpaired *t*-test) larger fluctuations with a latency of around 7 ms after each MN spike (green; $n = 7$ animals).

To determine the kinetics of myoplasmic calcium also at different MN firing frequencies, we stimulated a single MN electrically at different frequencies (0.5, 1, 2, 5, 10 Hz), spanning the normal range during flight, and imaged the resulting calcium signals in the respective DLM target fibre. Discrete

myoplasmic calcium signals were observed for the entire frequency range tested (Fig. 4c and Supplementary Video 4), although summation as caused by an incomplete return to resting calcium levels between MN spikes starts at 2 Hz and increases at higher frequencies. However, for firing frequency ranges of MN1–5 that are observed during normal flight (Fig. 1c), in each muscle fibre, myoplasmic calcium fluctuates around each corresponding MN spike. Given that we did not observe any differences in the temporal coupling of MN firing and myoplasmic calcium across different stimulation frequencies or different DLM fibres, these data can be extrapolated to all DLM fibres for splay state (Fig. 4d (top)) as compared to synchronous MN firing (Fig. 4d (bottom)). Given that each DLM fibre is innervated by one MN only (Fig. 1a), MN firing in splay state causes splayed-out calcium signals across the different muscle fibres (Fig. 4d (top)), each peaking around 6 ms after the respective MN spike and declining with a time constant of around 80 ms (Fig. 4b (middle)). The bottom trace shows the resulting average myoplasmic calcium across fibres (Fig. 4d (top)). By contrast, during synchronous firing of MN1–5, the calcium signals in all DLM fibres are time locked, which results in much larger calcium fluctuations across DLM fibres (Fig. 4d (bottom)). This extrapolation of the average myoplasmic calcium levels across the DLM fibres predicts eightfold smaller fluctuations in wingbeat frequency over time in normal splay state than for synchronous MN firing. We confirmed this prediction by averaging changes in wingbeat frequency around MN spikes for splayed-out (Fig. 4e (blue line)) as opposed to synchronous firing (Fig. 4e (green line)). Thus, the splay state serves to minimize wing power fluctuations.

Taken together, splay-state asynchronous flight motor patterns are conserved across individuals and species, are produced by a minimal CPG of weakly electrically coupled MNs, and serve constant wingbeat power output at a given power demand. This provides a comprehensive view of the asynchronous flight CPG network structure and the resulting functional consequences for one of the most abundant forms of locomotion on earth. Moreover, we provide a theoretical background for firing d esynchronization/synchronization in small gap-junctional networks in the context of their specific cellular excitability. The underlying mechanism is generic, predicting desynchronizing functions of electrical synapses beyond the motor system of insects. Electrical synapses can therefore be used for operations such as sign-reversal and ensemble firing desynchronization, a

functional versatility and impact on neural circuit dynamics that was previously attributed to chemical synapses alone.

Methods

Animals

D. melanogaster were reared in 68 ml transparent food vials (Kisker Biotech, 25 mm × 95 mm) filled with 10 ml of standard cornmeal/glucose/yeast/agar diet at 25 °C and 60% humidity under a 12 h–12 h light–dark cycle. For experiments, adult 2–5-day-old male flies were used. We used Canton-Special (Canton-S) as the wild-type control. To selectively target the DLM MNs (MN1–5) with UAS-driven transgenes we used a split GAL4 driver line that combines GMR23H06 (BDSC, 49050 discontinued) and GMR30A07 (BDSC, 49512) from the Rubin Collection[47]. The GMR23H06 enhancer expresses the activating domain (AD) and the enhancer GMR30A07 expresses the DNA-binding domain (DBD). Overlap of both expression patterns—and, therefore, functional GAL4 expression—is restricted to the MN1–5 (Extended Data Fig. 5a,b). The following transgenes were expressed in MN1–5 to manipulate gap junction expression and function. *shakB* RNAi (BDSC, 57706) targets all *shakB* isoforms and was used to KD gap junctions in MN1–5. For gain of function through overexpression of gap junction protein the shakB(N+16) isoform was used (gift from P. Phelan). KD of inhibitory synaptic transmission through GluCls or RDL GABA-ARs was conducted by targeting the expression of UAS-RNAi^{KK} for GluClα (VDRC, 105754) or UAS-Rdl-RNAi (VDRC, 41103) to MN1–5 with the DLM MN GAL4 driver GMR23H06. To test for electrical synapse function by genetic manipulation, UAS-RNAi-KD of *shakB* was compared with UAS-ShakB overexpression and empty UAS-control with 7 biological replicates in each group. To test for effects of GABA-ARs and GluClRs on flight patterns, UAS-RNAi-kd for each receptor type was expressed in MN1–5 and the effect on flight patterns was compared with the respective genetic controls, with 8 biological replicates for GluClRs and 10 replicate animals for GABA-ARs. For all genetic manipulations, the order of experiments was fully randomized and analysis was conducted without knowledge of the genotype. Other insect species were obtained from pet

shops as feed insects and *Apis mellifera* (at 4 to 6 weeks of age) from the local apiculture of the Johannes Gutenberg University Mainz.

For in vivo optogenetic manipulation of the activity of MN1–5 during flight, we expressed UAS-XXL channelrhodopsin (XXL-ChR) (BDSC, 58374)[48] either under the control of our DLM-MN-specific split GAL4 driver (split GMR23H06+GMR30A07) to enable light activation of MN1–5 (7 replicate animals) or, alternatively, in presynaptic cholinergic neurons under the control of Cha-GAL4 (11 replicate animals). Direct light activation of MN1–5 during flight was used to test whether the CPG correctly shapes MN1–5 firing into preferred sequences even when firing rates are artificially increased. Light activation of presynaptic cholinergic neurons during flight was used to test whether the CPG translates artificially presented, unpatterned synaptic input into patterned output from MN1–5. For in vivo calcium imaging in DLM fibres (6 replicate animals) the genetically encoded calcium indicator UAS-GCaMP8f[49] was expressed under the control of Act88-GAL4[50].

We used targeted expression of *FMRP*^{RNAi} in MN1–5 (*w;UAS-FMRP*^{RNAi}*/23H06-ADZ UAS-CD4-td-GFP;30A07-DBD/+*) to visualize weak electrical coupling by dye coupling. A reduction in FMRP is reported to increase both the uptake of small tracers into *Drosophila* neurons during iontophoresis and, consequently, also increase dye uptake into neurons that are electrically coupled through *shakB*-encoded gap junctions[28].

No ethical oversight was required for work with insect species. The work was according to the guidelines of research with transgenic invertebrate animals of the state of Rhineland-Palatinate in Germany.

Generation of UAS-Shab flies

Shab cDNA was provided by F. Sigworth[51] in pBluescript KSM. *Shab* was then PCR amplified with the following primers:

TTCAGGCGGCCGCGGCTCGAGaacttaaaaaaaaaaatACAA**ATG**GTCGGGCAATTGCAAGG (forward primer; the XhoI restriction site is underlined, the *Shab* start is bold and underlined) and

CCTTCACAAA**<u>GAT</u>**CC<u>TCTAGA</u>gactcactatagggcgaattgg (reverse primer; the XbaI restriction site is underlined, the *Shab* terminal sequence is bold and underlined). The primers included overhangs matching the pJFRC81 attB-site-directed expression vector (36432, Addgene). pJFRC81 was cut with XhoI and XbaI restriction enzymes according to the provider's protocol (New England Biolabs, R0146S, R145S) to remove the GFP sequence from the vector. Using NEBuilder according to provider's instructions (New England Biolabs, E5520S), the amplified *Shab* cDNA was inserted into the cut pJFRC81 vector, thereby replacing the GFP. The PCR-amplified *Shab* cDNA was not purified before assembly with NEBuilder. The assembled construct was then transformed into *Escherichia coli* Dh5α bacteria and streaked out onto ampicillin (50 μg ml^{-1}) containing LB media agar plates. Overnight colonies were picked and grown in ampicillin containing LB media (50 μg ml^{-1}). DNA was isolated using the Maxiprep kit (Qiagen, 12162). Using specific primers, we determined that the cDNA used was *Shab* splice variant K (Supplementary Table 5). The correct sequence was confirmed by sequencing. For generation of transgenic flies, the Shab-pJFRC81 attB vector was injected into fly embryos at the attP2 landing site by BestGene. For experiments, homozygous male w;;10xUAS-Shab-pJFRC81attP2 flies were crossed to female w;GMR23H06-ADZ;GMR30A07-DBD (DLM-split GAL4 driver) flies and male F_1 progeny were tested.

Electromyography/extracellular recording of MNs

The in vivo activity of MN1–5 during tethered flight was monitored by extracellular voltage recordings from their respective muscle fibre. One MN action potential causes one large, transient postsynaptic depolarization (Fig. 4b (top trace)) in its target DLM fibre (for innervation, see Fig. 1a). As each DLM fibre is innervated by one MN only, and each MN spike causes the same distinct evoked postsynaptic response, the EMG spikes recorded from the DLM fibres are a 1:1 reflection of the activity patterns of MN1–5.

To prepare flies for in vivo EMG recordings during tethered flight, animals were briefly cold-anaesthetized (20 s in an empty 68 ml plastic vial on ice), transferred dorsal side up onto a cold metal plate (~3 °C) and glued (clear glass adhesive (Duro; Pacer Technology)) with head and thorax to a triangle-

shaped tungsten hook (0.1 mm diameter). Curing of the glass adhesive was induced by exposure to ultraviolet light (Mega Physik Dental Cromalux-E Halogen Curing Light Unit) for 45 s. The flies were kept individually for 10 min to recover from the cold anaesthesia. The rested flies were then mounted in the set-up to a clamp attached to a micromanipulator. Tarsal contact to a small polystyrene bead prevented unwanted engagement into flight. Next, a light barrier, consisting of a red laser, an aperture to reduce beam width to approximately 1 mm and a LED light sensor, was positioned so that the light beam was broken by each wingbeat. After positioning of this laser-based wingbeat detector, tungsten electrodes for extracellular MN recordings were inserted. A reference electrode was inserted into the last abdominal segment. To record the activity of MN5 and MN4 from the dorsal-most DLM fibres, we inserted one sharpened tungsten wire into the dorsal thorax in front of the anterior dorsocentral bristle, making sure not to cross the midline of the thorax. This resulted in extracellular recordings of two units that could easily be distinguished by amplitude and shape. Confirmation of unit identity was achieved by adjusting the tungsten insertion depths and monitoring the resulting amplitude changes of the recorded units. As MN5 innervates the dorsal-most DLM fibres, it is the first unit to appear after shallow dorsal electrode insertion. Moving deeper lets a second unit appear that belongs to MN4, which innervates the fibre just beneath. This MN4 unit increases in amplitude with increased insertion depth in DLM4, while the amplitude of MN5 decreases. This procedure allowed unambiguous allocation of the two units to MN5 and MN4. To record up to 5 DLM units (MN1–5) simultaneously, additional tungsten electrodes were inserted. A second electrode was inserted four small bristles in front of the first electrode, which recorded MN4 and MN3. A third electrode was inserted anterior on the same line as electrodes 1 and 2 at the position where small bristles begin to appear on the thorax. Again, unit identification was conducted by altering electrode insertion depth at the beginning of each experiment. Starting from dorsal and slowly moving ventrally, the first unit to appear is MN5, the second one is MN4, the third one is MN3 and so on. Electrode depths were adjusted until the subsequent unit that appeared while moving deeper had approximately half of the amplitude of the first unit, so that both were easily distinguishable. In some recordings, it was also possible to pick up three clearly separable units simultaneously, but spike sorting was more time consuming.

All extracellular recordings were amplified at 1,000× using the AM-Systems 1700 extracellular amplifier. Electrophysiological and wingbeat recordings as well as a frame trigger signal from high-speed video (see below) were digitized with the Axon Digidata 1550B (Molecular Devices), acquired with AxoScope (v.10.7) and stored on a PC. Spike sorting was conducted offline with the spike sorting function in Spike2 (v.7.2). Data were further analysed using Spike2, and custom Python routines were created using Jupyter notebook to count pattern probabilities and create phase histograms. For additional functions, the Python libraries NumPy, pickle, SciPy, Matplotlib and seaborn were imported. All of the Python scripts are available on request. Additional data analysis and statistics were conducted using Microsoft Excel Professional Plus 2019 (v.2110) and GraphPad Prism (v.9.2.0).

All of the animals that were successfully recorded for at least 10 min of flight were included in the analysis of phase relationships. For spike processing, the DLM recordings (.abf format) were imported into Spike2 and sorted with the wavemark function, therefore creating templates for each unit. Template-matching spikes were visualized separately and synchronous spikes (the sum of the amplitude and waveform of two distinct MN units) were split into the two according units. The timepoints of the peaks of all marked spikes were then imported into single-event channels. The first and last 3 s of each flight bout were discarded to restrict the analyses to stable in-flight motor patterns. At the beginning of flight bouts, the MN spikes occur at higher frequency and with the highest irregularity, whereas, at the end of a flight bout, the firing frequency decreases. Phase histograms were created from the event channels. Each ISI of an event channel with the timepoints of the units of a given MN was compared to when the units of a second MN occur within the ISIs and counted into bins. All ISIs were normalized to 1 (phase 0 to phase 1), divided into 100 bins, the events of the second MN unit were counted into the respective bins and divided by the total number of events (relative occurrence; Extended Data Figs. 1, 2, 6 and 8). To create phase histograms with the phase 0 in the middle of the diagram (Fig. 2b and Extended Data Figs. 3b, 5e and 6e,f), the last 50 bins were copied in front of the first 50 bins and the *x* axis was changed to range from phase −0.5 to phase 0.5.

All insect species larger than *Drosophila* were cold anaesthetized for up to 20 min in the fridge at 5 °C and fixated in a 3D printed device to hold the insect in place. The .stl file is available on request. Instead of a rectangular tungsten hook, the wire is bent in a half circle and glued onto the outer circumference of the posterior dorsal region of the thorax to increase the adhesive surface.

High-speed video

In addition to electrophysiological recordings of MN activity and laser-based wing detection, selected times of tethered flight were simultaneously recorded as high-speed video using the Photron FASTCAM Mini UX100 camera and Photron FASTCAM Viewer software (PFV, v.3.6.9.0) at 5,000 fps. As illumination, two infrared lights (Sygonix IR illuminator with 48 LEDs) were used. Videos of *Drosophila* (Supplementary Video 1), goldfly (Supplementary Video 2) and honey bee (Supplementary Video 3) during tethered flight are provided.

Electrolytic sharpening of tungsten electrodes

The tips of tungsten rods (diameter of 100 µm for *Drosophila*, 125 µm for all larger insects) were sharpened electrolytically by repeated dipping in a $NaNO_2$ KOH solution (10.3 M $NaNO_2$ and 6.05 M KOH in double-distilled H_2O). Tungsten rods of 1 cm length were crimped onto metal rods. For this, the wire was placed onto the rod and covered with a 0.5 mm ferrule. A crimping tool compressed it to a pluggable tungsten wire electrode. Now the electrode was placed into an alligator clip, connected to a stimulator (Grass SD9 square pulse stimulator) and repeatedly dipped into the sharpening solution, which was connected to the other pole of the stimulator. A monophasic current with 100 Hz, 40 V and 1 ms duration was applied. The tip was frequently moved in and out of the solution to form a thin tip. Finally, electrodes were rinsed with distilled water.

Double patch-clamp recordings of MN pairs

After removing the legs and wings, 2–3-day-old adult *Drosophila* males were pinned into a Sylgard-coated lid of a 35 mm Petri dish and fixed

ventral side up with minute pins in the head and through the tip of the abdomen. After submerging in normal saline (128 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2$, 4 mM $MgCl_2$, 5 mM HEPES and ~35.5 mM sucrose, the pH was adjusted to 7.24 with 1 N NaOH, and osmolality was 300 mOsm kg^{-1}) the thoracical cuticle was removed with fine iris scissors to expose the VNC. The specimen was then rinsed thoroughly with saline. After mounting the preparation onto the stage of an upright Zeiss Axio Examiner epifluorescence microscope, the VNC was viewed with a ×40 water-immersion lens (Zeiss W Plan Apochromat ×40/1.0 NA, DIC VIS-R), the UAS-6xmCherry-expressing MN1–4 (genotype: *w;23H06-ADZ UAS-6xmCherry/+;30A07-DBD/+*) were viewed through a TRITC filter set. The ganglionic sheath and debris hampering access to MN1–4 somata were removed by repeated application of 1% protease type XIV (from *Streptomyces griseus*, Sigma-Aldrich, P5147) through a patch pipette with a manually broken tip that was then also used to remove loosened debris[52]. Before recording, the specimen was washed with saline for 5 min through a gravitation perfusion system at around 2 ml min^{-1}, the bath volume was around 300 µl. Recordings were performed using patch pipettes (borosilicate glass capillaries; outer diameter, 1.5 mm; inner diameter, 1 mm, without filament, World Precision Instruments, PG52151-4) pulled with a PC-10 vertical puller (Narishige) filled with internal patch solution (140 mM K-gluconate, 2 mM Mg-ATP, 2 mM $MgCl_2$, 11 mM EGTA, 10 mM HEPES. pH was adjusted to 7.24 with 1 N KOH, osmolality was adjusted to 300 mOsm kg^{-1} with glucose if necessary) that approached the two MNs from opposite sides. Patch pipette tip resistance was between 5 and 6 MΩ with these solutions. Double patch recordings were performed by connecting each patch pipette to a separate Axopatch 200B amplifier (Molecular Devices). Data were filtered at 5 kHz through a lowpass Bessel filter, digitized through an analogue/digital converter (Digidata, 1440), and signals were recorded using pClamp10.7 software (both Molecular Devices). Output gain was 10×. For double recordings, one MN was approached with a patch pipette. After giga seal formation, pipette capacitance artifacts were cancelled manually and whole-cell configuration was established at a holding potential of −70 mV. After setting whole-cell capacitance compensation, correction and prediction values as well as series resistance compensation using the respective dials of the amplifier (only used to judge on recording quality as all of these compensations are disabled in current clamp mode), the

recording was disconnected from external influence by using the I=0 setting of the amplifier, making it possible to revert to bath mode without disturbing the already established recording of the first MN. Then, recording of the second MN was established identically. To monitor both recordings, separate channels of the digitizer and in the pClamp10.7 software were used. Quality parameters were as follows: giga seal, >5 GΩ; membrane potential of −70 mV was held with a holding current smaller than ±100 pA; series resistances of >15 MΩ were not accepted to ensure good control when applying current injections; only if these criteria were met for both MNs, the recordings were switched to current clamp mode. Resting membrane potential was around −60 mV without current injection. To determine coupling strength and rectification parameters of gap junctions between MNs, depolarizing and hyperpolarizing current was injected into one MN while the other was monitored simultaneously, and vice versa. Slow tonic firing was induced in one or both MNs at rates of between 3 and 8 Hz, as is observed during flight behaviour, by small somatic current injections while monitoring the respective other MN. For input–output relationships, firing was induced by 1,000 ms square pulse current injections up to 1 nA in 0.1 nA increments.

Intracellular dye filling

Intracellular dye fills with neurobiotin were used to detect dye coupling between electrically coupled MNs. For intracellular dye fills, *FMRP^{RNAi}* was targeted to MN1–5 (*w;UAS-FMRP^{RNAi}/23H06-ADZ UAS-CD4-td-GFP;30A07-DBD/+*), because reductions in FMRP have been shown to increase both neurobiotin uptake into *Drosophila* neurons during iontophoresis and, consequently, also increased dye uptake into neurons that are electrically coupled through *shakB*-encoded gap junctions[28]. After removing the legs and wings, 2–3-day-old adult male *Drosophila* were pinned in a Sylgard-coated lid of a 35 mm Petri dish and fixed dorsal side up with two minute pins, one through the head and one through the abdomen. After submerging the specimen in normal saline (128 mM NaCl, 2 mM KCl, 1.8 mM $CaCl_2$, 4 mM $MgCl_2$, 5 mM HEPES and ~35.5 mM sucrose, pH was adjusted to 7.24 with 1 N NaOH, and osmolality was 300 mOsm kg^{-1}), it was opened along the dorsal midline up to the neck connectives with iris scissors. The cut DLM was then pinned to the sides with one minute pin each to expose the gut, inner organs and VNC. The gut, salivary glands as

well as other inner organs were removed to fully expose the VNC. The specimen was then rinsed thoroughly with saline to remove excess debris. After mounting the preparation onto the stage of an upright Zeiss Axio Examiner epifluorescence microscope, the VNC was viewed using a ×40 water-immersion lens (Zeiss W Plan Apochromat 40x NA 1.0, DIC VIS-R), the UAS-CD4-td-GFP-expressing MN1–5 were viewed through a FITC filter set. The ganglionic sheath was removed focally using a broken patch pipette filled with 1% protease type XIV (from *Streptomyces griseus*, Sigma-Aldrich, P5147) to allow access to MN somata[52]. Dye fills were performed using sharp glass microelectrodes pulled from filamented borosilicate glass capillaries (Sutter BF100-50-10; tip resistance, ~40 MΩ) with a Sutter P-97 Flaming Brown microelectrode puller. The tip was filled with a 50/50 mixture of TRITC-Dextran 3000 lysin fixable (Invitrogen, 3308) and neurobiotin (Vector Labs, SP-1120) dissolved in 2 M potassium acetate, and the shaft was then filled with 2 M potassium acetate leaving an air bubble between the dye-filled tip and the shaft to avoid dilution of the dye. MNs were impaled by a short buzz (which makes the electrode tip vibrate for a set amount of time, here around 40 ms) with a remote buzz connected to an Axoclamp 2B intracellular amplifier (Molecular Device) that was also used for dye filling. MNs were filled iontophoretically by positive current injection (between 0.5 and 1 nA) in bridge mode until the MN was judged to be filled, after around 10 min. The microelectrode was then removed, and the specimen was fixed with 4% paraformaldehyde at room temperature for 50 min, without shaking. This was followed by three rinses with PBS and three washes for 20 min PBS and six washes for 20 min with 0.5% PBS-Triton X-100 (PBT, Triton X-100, Sigma-Aldrich, T8787), with shaking. The preparation was then incubated with Streptavidin coupled to Alexa 647 (Thermo Fisher Scientific, S-21374) in 0.3% PBT at room temperature in the dark for 2 h with shaking. Streptavidin Alexa 647 was then removed, the preparation was rinsed several times with PBS and then washed three times for 20 min with PBS at room temperature in the dark with shaking. This was followed by an ascending ethanol series: 50%, 70%, 90%, 100% for 10 min each. The preparation was then mounted in methylsalicylate, topped with a high-precision (170 ± 5 µm) cover slip and sealed with clear nail polish. The dye fill was then visualized using the Leica TCS SP8 confocal laser microscope with a helium–neon laser. Alexa 647 was excited at 633 nm and emission was detected between 650 and 680 nm

with a photomultiplier tube. Images were taken with a ×40 oil objective (NA 1.3) with a 1.75 digital zoom at a resolution of 1,024 × 1,024 pixels and a *z*-step size of 1 µm.

Optogenetics

For optogenetic stimulation, UAS-channelrhodopsin was either expressed under the control of a DLM-MN-specific Split-GAL4 line (GMR23H06-ADZ attP49; GMR30A07-DBD attP2; see Extended Data Fig. 5a,b for expression patterns specifically in MN1–5), or under the control of ChaT-GAL4 in cholinergic neurons. We tested multiple different optogenetic transgenes. UAS-CsChrimson for red light activation caused few MN spikes followed by silence, probably caused by a depolarization block, and was therefore not suitable for optogenetic stimulation during ongoing flight. By contrast, blue light stimulation of channelrhodopsin XXL reliably increased MN firing frequencies during flight, and therefore was the most effective transgene for optogenetic activation experiments during flight. As a control, the blue light stimulation was applied in an identical manner to flies without expression of UAS-channelrhodopsin XXL (DLM-MN split-GAL4 and ChaT-GAL4 crossed to w1118 flies, respectively). Blue light stimulation alone had no effect on MN firing frequencies during flight. The light stimulus was applied though fibre optics directed to the ventral thorax from underneath to avoid stimulation of cholinergic neurons in the brain. The light stimulus was presented constantly so that optogenetic activation of either cholinergic pre-motor neurons (Extended Data Fig. 3) or MN1–5 (Extended Data Fig. 5c,d) was not patterned. Moreover, channelrhodopsin XXL activates within milliseconds after blue light stimulation but has very slow inactivation kinetics with an off time constant of 76 ± 12 s (ref. [48]). Thus, increased firing rates during flight outlast the blue light stimulus (Extended Data Figs. 3 and 5c).

In vivo calcium imaging from DLM fibres during flight

For in vivo calcium imaging in DLM fibres, the genetically encoded calcium indicator UAS-GCaMP8f[49] was expressed in muscle under the control of Act88-GAL4[50]. All earlier GCaMP versions exhibit signal decay kinetics that are too slow to measure the time course of decay of myoplasmic

calcium signals as induced by a MN action potential. Calcium imaging was conducted using a Hamamatsu CMOS (C11440-42U) camera mounted to a Zeiss Axioscope 2FS fluorescence microscope equipped with a ×10 lens that allowed enough working distance for wingbeats under the microscope. Images were acquired at a frame rate of 98 Hz with 16 bit image depth at a 512 × 512 pixel spatial resolution. Myoplasmic calcium responses in DLM fibres were imaged in vivo through the dorsal cuticle either during tethered flight, or after electrical stimulation of selected MNs (Supplementary Video 4 shows myoplasmic calcium responses of DLM fibre 6 in response to electrical stimulation of the MN5 that innervates DLM fibres 5 and 6; Fig. 1a). Myoplasmic calcium signals were acquired using Hamamatsu photonics software (TOKUPIC v.1.0) and presented as percentage change ($\Delta F/F$) as previously described[21].

Software

Figures were created in Corel DRAW 2021 and Adobe Illustrator 2021. Acquisition for electrophysiology was conducted with pClamp (v.10.7) and data were further analysed using Spike2 (v.7.2) and GraphPad Prism (v.9.2.0). Calcium imaging data were acquired using Hamamatsu photonics software (TOKUPIC v.1.0). Model simulations were performed using the brian2 package (version 2.4.2.) for python[53]. The bifurcation analysis of the model was carried out using AUTO-07P[54]. Statistical tests were performed using scipy.stats of SciPy (v.1.7.1).

Computational modelling and equations

Conductance-based MN model and gap junctional circuit model

Individual MNs are described by a single-compartment model with spike generating Na^+ and K^+ currents, I_{Na} and I_K, based on ref. [39] and further refined on the basis of in situ current clamp and voltage clamp recordings from our laboratory (see below). For $I_{Na,}$ the half-activation voltage of the activation gate is −33 mV, the half-activation voltage of the inactivation gate is −39.14 mV and the reversal potential E_{Na} is 55 mV. For I_K (referred to as I_{shab} in the equations below) the half-activation voltage is −42.14 mV and

the reversal potential E_K is −72 mV. The maximal sodium conductance g_{Na} is 0.4312 µS and the maximal potassium conductance g_{shab} is varied between 0.13768216 µS and 0.34496 µS throughout the paper (see the 'Shab-induced bifurcations' section; a list of single-neuron-model parameters is provided in Supplementary Table 2). For the gap junctional circuit model, identical single-neuron models were coupled by linear non-rectifying gap junction currents, I_{gap}, with properties based on dual patch current clamp recordings from MN pairs. The current–balance equation of the i^{th} neuron reads

$${C}_{{\mathtt{m}}}{\dot{v}}_{i}={I}_{{\rm{i}}{\rm{n}}}-{I}_{{\mathtt{L}}}({v}_{i})-{I}_{{\rm{s}}{\rm{h}}{\rm{a}}{\rm{b}}}({v}_{i})-{I}_{{\rm{N}}{\rm{a}}}({v}_{i})+\sum _{j\ne i}{I}_{{\rm{g}}{\rm{a}}{\rm{p}}}^{ij}({v}_{i},{v}_{j}).$$

where v is the membrane voltage and C_m is the membrane capacitance. The leak current I_L with the leak conductance g_L, the potassium current I_{shab}, and the sodium current I_{Na} are defined as

$${I}_{{\mathtt{L}}}(v)={g}_{{\mathtt{L}}}(v-{E}_{{\rm{L}}}),$$

$${I}_{{\rm{s}}{\rm{h}}{\rm{a}}{\rm{b}}}(v)={g}_{{\rm{s}}{\rm{h}}{\rm{a}}{\rm{b}}}{b}^{4}(v-{E}_{{\mathtt{K}}}),$$

$${I}_{{\rm{Na}}}(v)={g}_{{\rm{Na}}}{m}_{{\rm{\infty }}}^{3}(v)(1-h)(v-{E}_{{\rm{Na}}}).$$

Gates of the voltage-dependent ion channels follow first-order kinetics

$${\tau }_{b}(v)\dot{b}={b}_{{\rm{\infty }}}(v)-b,$$

$${\tau }_{h}(v)\dot{h}={h}_{{\rm{\infty }}}(v)-h,$$

The bidirectional, non-rectifying gap junctions with linear charge transfer (Fig. 2e,f) are described by the following current

$${I}_{{\rm{gap}}}^{ij}\left({v}_{i},{v}_{j}\right)={g}_{{\rm{gap}}}^{ij}\left({v}_{j}-{v}_{i}\right).$$

(1)

Parameters common to all simulations are provided in Supplementary Table 2. For simulations with homogeneous coupling (Fig. 3b–d,j (left)) a coupling strength of ${g}_{{\rm{gap}}}^{ij}\,=\,43.5\,{\rm{pS}}$ for all $i \neq j$ was used. To keep the cases as comparable as possible, the mean coupling strength in the simulations with heterogeneous coupling (Fig. 3h,j) was chosen to be the same as in the homogeneous case. The ratios of coupling coefficients for MN1–4 were chosen to be the same as measured in vivo (Fig. 2f). MN5 was coupled in with a coupling strength half of the mean of the other connections. This yielded ${g}_{{\rm{gap}}}^{12}={g}_{{\rm{gap}}}^{21}={g}_{{\rm{gap}}}^{34}={g}_{{\rm{gap}}}^{43}=86.59\,{\rm{pS}}$, for the MN5 connections ${g}_{{\rm{gap}}}^{15}={g}_{{\rm{gap}}}^{51}={g}_{{\rm{gap}}}^{25}={g}_{{\rm{gap}}}^{52}={g}_{{\rm{gap}}}^{35}={g}_{{\rm{gap}}}^{53}=\,{g}_{{\rm{gap}}}^{45}={g}_{{\rm{gap}}}^{54}\,=$ 27.19 pS and 38.27 pS for the remaining connections. The coupling strength in the case of strong coupling (Fig. 3d) was chosen to be 3 nS for all connections.

Stochastic and deterministic simulations of the model were performed using the brian2 package (v.2.4.2) for Python[53]. For stochastic simulations, the Heun method was used with a time step of 3 µs. For noisy simulations, a zero-mean white-noise current was added to the current balance equation (using the brian2 xi variable) with a noise strength of $\sigma =0.949\,{\rm{pA}}\sqrt{{\rm{ms}}}$ (Fig. 3d,h,j). For the deterministic simulations in Fig. 3b,c, the classical Runge–Kutta method (RK4) was used with a time step of 100 µs.

In addition, Q represents $\frac{e}{{k}_{{\rm{B}}}T}$, which is the elementary charge e divided by the Boltzmann constant k_B and the temperature T in Kelvin. Throughout the simulations, Q was approximated as $39.2/V$.

The model parameters common to all models are summarized in Supplementary Table 2; the activation curves of the gates and the kinetic time scales are provided in Supplementary Table 3. Some parameters and equations were reported inconsistently in the original publication. Because the authors[39] provided their code, we were able to extract the parameters

used for their simulations, which are reported here. These updated parameters were validated using additional current-clamp and voltage-clamp recordings from our laboratory. Furthermore, to produce different onset bifurcations, the Shab conductance was changed (see below) and the input current was adjusted to move the system close to onset. The values differing between models are summarized in Supplementary Table 4.

For almost all simulations, neurons were initiated in random phases, by drawing timepoints from within one ISI and initializing the simulation with the corresponding state vector of a periodically spiking uncoupled neuron. For the simulations in Fig. 3c, the same initial conditions were used for each coupling strength. For the example in Fig. 3b, the initial conditions were not random, but hand-picked to yield a sequence that matches one of the preferred sequences in Fig. 1f.

Gap junctions that adhere to equation (1) affect the intrinsic properties of neurons because they act as effective leak and capacitance changes. For example, gap-junction currents can influence firing frequencies depending on the synchronization state of the network. In Fig. 3c, the coupling is homogeneous \({g}_{{\rm{gap}}}^{i,j}={g}_{{\rm{gap}}}\) for all $i \neq j$ between pairs but the coupling strength $g_{\rm gap}$ is varied. The effect of coupling on firing frequencies is strongest in the intermediate coupling range. In the weak coupling range, the effect on frequency is small because the current flow through gap junctions is small. In the strong coupling range, the neurons synchronize and, consequently, almost no gap-junctional current flow occurs. As frequency homeostasis is not part of this model, we accept biologically implausible firing frequencies in the intermediate range. In biological neurons, a wide range of mechanisms—for example, based on additional ion channels, energy consumption or ionic concentrations—could stabilize the frequencies[55].

Shab-induced bifurcations

A local bifurcation analysis of the fixpoints in the model was shown previously[39]. The important quantity for predicting network states of the CPG is the PRC (Fig. 3e), which is closely related to the spike-onset bifurcation of the neuron and, therefore, the neuron's excitability class. The

bifurcation diagram in Extended Data Fig. 9a adds non-local and codim-2 bifurcations that organize transitions between excitability classes and their PRCs such as big SNL and small SNL[37,56]. Depending on the Shab channel density, g_{shab} on the ordinate, different onset bifurcations that change the resting state into the spiking mode occur. In the middle range of g_{shab} levels, spiking commences through a well-known SNIC bifurcation. The SNIC region is encapsulated by two SNL points. At the lower end, for g_{shab} smaller than the small SNL (sSNL) point, spiking commences through a small homoclinic loop (HOM, green line in Extended Data Fig. 9a) at lower input currents, I_{in}, than the saddle-node bifurcation (blue line in Extended Data Fig. 9a). At the upper end of the SNIC range, a series of bifurcations is traversed that eventually leads to spike onset through a fold of limit cycles (FLC, bifurcation line not shown) and a subsequent subcritical Hopf bifurcation, here termed the Hopf regime for simplicity. Bifurcations leading from SNIC to the Hopf excitability class are the big SNL, the neutral saddle loop (not shown) and also a cusp[56]. For simplicity, in the main text, when using the term SNL, we refer to sSNL. The Bogdanov–Takens point in between creates an Adronov–Hopf line (red line in Extended Data Fig. 9a) that eventually turns back to create the excitation block at higher input levels.

The consequences for the phase relationships of coupled neurons can be understood from the neuron's PRC at different bifurcations (Extended Data Fig. 9a). A neuron's ability to produce arbitrarily low firing rates depends on its excitability class (determined by the spike onset bifurcation). So does the PRC shape[57,58,59]. The insets in Extended Data Fig. 9a show prototypical PRCs for the three onset bifurcations, which can be achieved at different levels of Shab-channel density. The exact parameters are summarized in Supplementary Table 4. A critical transition with repercussions for the CPG network states is the symmetry breaking in the PRC at the sSNL bifurcation as described in the next section[59].

An FLC/Hopf onset bifurcation as spike onset can be ruled out as a model for the biological MN1–5 network based on the following grounds: the region in which tonic spiking commences through the FLC/Hopf bifurcation has a finite, non-zero frequency at the rheobase. For the present model (Extended Data Fig. 9a (top arrow)), this is at around 47 Hz and therefore

above the dynamic range of the in vivo measures *f–I* curves from the MNs, which operates between 0–40 Hz (3—15 Hz during normal flight; Fig. 1e).

The bifurcation analysis was performed using AUTO-07P[54].

Mechanism of splay state generation

Our theoretical analysis predicts that dynamics close to the SNL point favour splay states[59]. This prediction is based on coupling functions, which are obtained from the biophysical conductance-based models using a phase-reduction (see the next section).

The coupling function $G(\psi)$ combines a neuron's PRC and a synaptic transfer term, which, for the system at hand, captures the voltage perturbations caused by gap-junctional coupling (equation (4)). The coupling function describes how one neuron's phase is shifted based on the phase difference to another neuron. Of particular interest is the odd part of the coupling function, $G^{\mathrm{odd}}(\psi) = G(\psi) - G(-\psi)$. It reflects the reciprocity of coupling between two neurons and extracts how coupling affects the phase difference in the cells' firing. Phase relationships of tonically firing neurons can be read from G^{odd} (Fig. 3e (bottom)).

Crucially, stable phase relationships depend on a stable fix point in this odd part of the coupling function, that is, a zero crossing with a negative slope. The shape of the coupling function in turn depends on the bifurcation type of the neuron (Extended Data Fig. 9a). For example, for a PRC of a neuron with SNIC dynamics, a small phase distance, $\Delta\varphi$, (neuron 1 fires shortly before neuron 2) is further decreased, whereas a large $\Delta\varphi$ (neuron 2 fires shortly before neuron 1) is further increased. This results in a stable fixpoint at phase 0, so that the neurons show synchronized firing (Fig. 3e (bottom)). By contrast, with a PRC at the SNL point, a small $\Delta\varphi$ is increased, whereas a large $\Delta\varphi$ is decreased. This results in a stable fixpoint at $\Delta\varphi = 0.5$ for each coupled neuron pair. For a network of >2 electrically coupled neurons, this causes a frustrated state because antiphase locking at $\Delta\varphi = 0.5$ cannot be achieved for all units at the same time. In a small network, such as the MN1–5 network with its five coupled neurons, the splay state represents a low-frustration solution that determines the most likely network state[60].

These analytical predictions based on coupling functions were corroborated by simulations of the full biophysical model. Near the SNL point, networks of five gap-junction-coupled neurons indeed showed a splay state (Fig. 3b), whereas neurons with a SNIC onset synchronized instead (Fig. 3h). These simulations show that, for our parameters, the phase reduction is valid and that coupling functions based on the phase reduction accurately capture the network's synchronization behaviour. On the basis of both coupling functions and simulations, we predict that the combination of coupling and intrinsic excitability determines the splay state in the biological MN1–5 network.

Phase-reduced circuit model with repulsive coupling

To generate coupling functions, we reduce the biophysical, conductance-based neurons to approximate phase oscillators[61]. This is warranted because the cholinergic inputs drive MN1–5 into a tonically spiking regime. The i^{th} phase oscillator is defined by its intrinsic, mean firing rate, f_i, and its phase response curve, $Z(\varphi_i)$, obtained from the biophysical model (Extended Data Fig. 9a). The PRCs are calculated based on a direct perturbation approach[62,63]. The network equation reads:

$${\dot{{\phi }}}_{i}={f}_{i}+\sum _{j\ne i}Z({{\phi }}_{i})g({{\phi }}_{i},{\phi }_{j}),$$
(2)

where $g(\varphi_i,\varphi_j)$ is the phase-dependent perturbation received through the gap junctions. It depends on the phases of the presynaptic and postsynaptic neurons. If the spike waveforms are uniform, the interaction term, based on equation (1) reads

$$g({{\phi }}_{i},{{\phi }}_{j})={g}_{{\rm{g}}{\rm{a}}{\rm{p}}}(v({{\phi }}_{j})-v({{\phi }}_{i}))/{C}_{{\rm{m}}}.$$

To understand the influence of the PRC on phase relationships in small networks, first consider the phase difference between two coupled neurons i and j, isolated from the network. Define their phase difference $\psi = \varphi_i - \varphi_j$.

Then, using approximate averaging theory, the slow evolution of the phase difference simplifies[33] to

$$\dot{\psi }=\nu +{G}^{{\rm{odd}}}\left(\psi \right).$$
(3)

Here, the small frequency detuning is $\nu = f_i - f_j$ and the averaged coupling function is defined as

$$G(\psi)={\int }_{0}^{1}Z({{\phi }}_{i})g({{\phi }}_{i},{{\phi }}_{i}-\psi)d{{\phi }}_{i}.$$
(4)

Stable fixpoints of equation (3) determine constant phase relations for neurons with similar intrinsic frequencies as illustrated in Extended Data Fig. 9b,c. In Fig. 3e the odd part, $G^{\rm{odd}}(\psi)$, and the fixpoints are shown for different choices of $Z(\varphi)$. The stable in-phase fixpoint directly translates to a synchronous network state also for network sizes of $N > 2$. However, this is not observed in the CPG recordings (Fig. 1a). For the network to show frustration, pairs of neurons must be phase-repellent.

The top spike trains in Extended Data Fig. 9b illustrate an example that generates stable antiphase fixpoints. To explain stable phase relationships graphically, the model was simplified, such that the shape of the coupling is a scaled version of the PRC (corresponding to gap junction coupling with delta spikes). Moreover, the mean PRC was absorbed into the average firing rate. The stability can be analysed using the iterated map of the phase differences. The return map reads $\psi_{n+1} = \psi_n + Z^{\rm{odd}}(\psi_n)$. Extended Data Fig. 9c uses iterated maps to show how a PRC with a negative slope at $\varphi = 0.5$ leads to a stable fixpoint in antiphase (left). An example with a positive slope at $\varphi = 0.5$ is shown in Extended Data Fig. 9b (bottom) and Extended Data Fig. 9c (right) and shows in-phase synchronization.

In conclusion, reciprocally, yet weakly coupled neurons, such as those connected through non-rectifying gap junctions, require an asymmetric coupling function (that is, with a substantial odd part) to show stable phase locking. For antiphase synchronization to be stable and the network in equation (2) to show frustration, the odd part of the coupling function must

have a stable fixpoint at phase 0.5. In the low-frequency limit, this can be achieved with a PRC that has a negative slope at $\varphi = 0.5$, such as the PRC at the SNL point.

Calculating model coupling coefficients

Coupling coefficients for the model were calculated analytically. For this, we looked at pairs of gap-junction-coupled neurons (not the whole network of five neurons) and it was assumed that the voltage is sufficiently negative for all voltage-gated ion channels to be closed. In this case, the relationship between coupling coefficient (CC) and gap junction strength (g_{gap}) reads

$${\rm{C}}{\rm{C}}=\frac{{g}_{{\rm{g}}{\rm{a}}{\rm{p}}}}{{g}_{{\rm{g}}{\rm{a}}{\rm{p}}}+{g}_{{\mathtt{L}}}}.$$

This relationship was verified in simulations that mimicked the experimental approach to determine coupling coefficients[29]. Specifically, two neurons coupled through gap junctions received steady current inputs below the firing threshold. The first neuron, n_1, received a baseline current $I_{\mathrm{base}} = -150$ pA, the other neuron, n_2, received an additional perturbation current $I_{\mathrm{base}} + \Delta I$ (ΔI from −5 pA to 5 pA). A third, uncoupled neuron n_3 was simulated with I_{base} input. Simulations were run until all neurons reached steady state, resulting in steady state voltages v_1, v_2 and v_3. Then, the coupling coefficients were calculated as ${\rm{CC}}=\frac{{v}_{1}-{v}_{3}}{{v}_{2}-{v}_{3}}$. Here, a low I_{base} ensures that the resulting CC is steady over the range of ΔI currents and matches the analytical results.

The splayness index

When spike times of all N neurons in a network are available, splayness can be quantified by comparing the neuron's phase differences to those that would arise in a perfect splay state.

To this end, we first discretized the time from the first spike of the last neuron that started spiking until the last spike of the first neuron that stopped spiking with a time step of 1 ms. If there were breaks in the recording (where

the fly stopped and started flying again), those times were omitted. Again, the breaks were defined from the last spike of the first neuron that stopped spiking until the first spike of the last neuron that started spiking. This way, all phases of all N neurons were sampled with a 1 ms time step at K time points $\tau_k = k$ ms for $k = 0, 1, ..., K - 1$.

For each neuron i, its phases were interpolated between its spikes $t_{n,i}$ as:

$${{\phi }}_{k,i}=\frac{{\tau }_{k}-{t}_{n,i}}{{t}_{n+1,i}-{t}_{n,i}},({t}_{n,i}\le {\tau }_{k} < {t}_{n+1,i})\,{\rm{f}}{\rm{o}}{\rm{r}}\,i=1,2,\ldots \,N.$$
(5)

Then, for each point in time, the phases were ordered such that

$${{\phi }}_{k,l} > {{\phi }}_{k,m},(l > m).$$

Phase differences were computed as

$${\psi }_{k,i}={{{\phi }}_{k,i+1}-{\phi }}_{k,i}\,{\rm{f}}{\rm{o}}{\rm{r}}\,i=1,...N-1$$

and

$${\psi }_{k,N}=1-\mathop{\sum }\limits_{i=1}^{N-1}{\psi }_{k,i}.$$

To calculate the splayness, these phase differences were compared to the phase differences of the most splayed state (splay state) and to those of the least splayed state (sync state):

$${\gamma }_{k}=\frac{\mathop{\sum }\limits_{i=1}^{N}{\left({\psi }_{k,i}-\frac{1}{N}\right)}^{2}}{(N-1){N}^{-2}+{(1-{N}^{-1})}^{2}}=\frac{N}{N-1}\mathop{\sum }\limits_{i=1}^{N}{\left({\psi }_{k,i}-\frac{1}{N}\right)}^{2}.$$
(6)

From this we constructed the time-averaged splayness

$$s=1-\sqrt{\frac{1}{K}\sum _{k}{\gamma }_{k}}.$$

(7)

With this definition, a perfectly splayed network yields $s = 1$ and a perfectly in-phase synchronized network yields $s = 0$.

Analysing pairs of neurons (synchronization index and phase histogram dips)

Many experimental recordings include only two of the five MNs (Fig. 3d,h). In these cases, the splayness index cannot be used. Here we relied on a pairwise synchronization measure instead. Specifically, we interpolated the phases of the two neurons in question according to equation (5). Then, a classical Kuramoto order parameter was evaluated:

$${r}_{k}=| \frac{1}{2}\mathop{\sum }\limits_{j=1}^{2}{{\rm{e}}}^{i2{\rm{\pi }}{{\phi }}_{k,j}}| .$$

Its time average $r=\frac{1}{K}{\sum }_{k}{r}_{k}$ measures in-phase synchronization of the neuron pair, where K is the number of time points.

A low synchronization index does not necessarily imply a splay state, as there are other desynchronized network states with equally low synchronization values. We therefore investigated alternative ways to gain information about splayness in the full MN1–5 network when in vivo recordings only comprised two neurons. The shape of pairwise phase histograms provides such information. Specifically, we observed that networks with high splayness often display a dip in pairwise histograms around zero. To test whether this dip is a good indicator for splayness, additional simulations were performed. With increasing noise strength, the dip around zero decreases, but it is more persistent than other histogram characteristics of splayed-out networks (Extended Data Fig. 6a). Furthermore, there was a strong negative correlation between the fraction of the MN4–MN5 pairwise histogram within (−0.1, 0.1) and the full-network splayness in experimental recordings for which all five neurons were available (Extended Data Fig. 6d). Together, this led us to conclude that the dip around zero is an appropriate indicator for splayness.

Comparing synchronization indices

To compare synchronization indices (Fig. 3d,h), Mann–Whitney *U*-tests were performed using scipy.stats of SciPy v.1.7.1. Tests were uncorrected

更多杂志请加vx
几百种杂志持续更新

because only a few comparisons based on our hypotheses were performed.

Regarding the results of Fig. 3d, MN4–MN5 synchronization indices for simulations with SNL onset and strong coupling (Mdn = 1.0) were higher than those for simulations with SNL onset and weak coupling (Mdn = 0.54). A two-sided Mann–Whitney *U*-test indicated that this difference was statistically significant, $U(N_{\text{strong SNL}} = 10, N_{\text{weak SNL}} = 10) = 100$, $P = 0.0002$. MN4–MN5 synchronization indices for flies with *shakB* RNAi (Mdn = 0.61) were higher than those for control animals (Mdn = 0.56). A two-sided Mann–Whitney *U*-test indicated that this difference was statistically significant, $U(N_{\text{shakB-RNAi}} = 9, N_{\text{control}} = 11) = 78$, $P = 0.0334$. MN4–MN5 synchronization indices for flies with *shakB* overexpression (Mdn = 0.84) were higher than those for control animals (Mdn = 0.56). A two-sided Mann–Whitney *U*-test indicated that this difference was statistically significant, $U(N_{\text{shakB-OE}} = 7, N_{\text{control}} = 11) = 73$, $P = 0.0008$.

Regarding the results of Fig. 3h, MN3–MN4 synchronization indices for SNIC simulations (Mdn = 0.99) were higher than those for SNL simulations (Mdn = 0.26). A two-sided Mann–Whitney *U*-test indicated that this difference was statistically significant, $U(N_{\text{SNIC}} = 10, N_{\text{SNL}} = 10) = 100$, $P = 0.0002$. MN3–MN4 synchronization indices for flies with *Shab* overexpression (Mdn = 0.52) were higher than those for control flies (Mdn = 0.43). A two-sided Mann–Whitney *U*-test indicated that this difference was statistically significant, $U(N_{\text{shab-OE}} = 10, N_{\text{control}} = 8) = 63$, $P = 0.0434$.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The electrophysiology and imaging data have been deposited at the Zenodo open data repository (https://doi.org/10.5281/zenodo.7737730). Computational modelling data and additional computational analyses of in

vivo electrophysiological data have been deposited at the Zenodo open data repository (https://doi.org/10.5281/zenodo.7740678).

Code availability

Code for implementing the computational model and for electrophysiological data analyses is available at Zenodo (https://doi.org/10.5281/zenodo.7740678).

References

1. Gordon, S. & Dickinson, M. H. Role of calcium in the regulation of mechanical power in insect flight. *Proc. Natl Acad. Sci. USA* **103**, 4311–4315 (2006).

2. Frye, M. et al. Molecular dynamics of cyclically contracting insect flight muscle in vivo. *Nature* **433**, 330–334 (2005).

3. Iwamoto, H. & Yagi, N. The molecular trigger for high-speed wing beats in a bee. *Science* **341**, 1243–1246 (2013).

4. Fry, S. N., Sayaman, R. & Dickinson, M. H. The aerodynamics of free-flight maneuvers in *Drosophila*. *Science* **300**, 495–498 (2003).

5. Farisenkov, S. E. et al. Novel flight style and light wings boost flight performance of tiny beetles. *Nature* **602**, 96–100 (2022).

6. Stork, N. E. How many species of insects and other terrestrial arthropods are there on Earth? *Annu. Rev. Entomol.* **63**, 31–45 (2018).

7. Deora, T., Gundiah, N. & Sane, S. P. Mechanics of the thorax in flies. *J. Exp. Biol.* **220**, 1382–1395 (2017).

8. Josephson, R. K., Malamud, J. G. & Stokes, D. R. Asynchronous muscle: a primer. *J. Exp. Biol.* **203**, 2713–2722 (2000).

9. Ikeda, J. H. & Koenig, J. H. Morphological identification of the motor neurons innervating the dorsal longitudinal flight muscle of *Drosophila melanogaster*. *J. Comp. Neurol.* **273**, 436–444 (1988).

10. Coggshall, J. C. Neurons associated with the dorsal longitudinal flight muscles of *Drosophila melanogaster*. *J. Comp. Neurol.* **177**, 707–720 (1978).

11. Consoulas, C., Restifo, L. L. & Levine, R. B. Dendritic remodeling and growth of motoneurons during metamorphosis of *Drosophila melanogaster*. *J. Neurosci.* **22**, 4906–4917 (2002).

12. Tyrer, N. M. & Altman, J. S. Motor and sensory flight neurons in a locust demonstrated using cobalt chloride. *J. Comp. Neurol.* **157**, 117–138 (1974).

13. Duch, C., Bayline, R. J. & Levine, R. B. Postembryonic development of the dorsal longitudinal flight muscle and its innervation in *Manduca sexta*. *J. Comp. Neurol.* **422**, 1–17 (2000).

14. Schlurmann, M. & Hausen, K. Motoneurons of the flight power muscles of the blowfly *Calliphora erythrocephala*: structures and mutual dye coupling. *J. Comp. Neurol.* **500**, 448–464 (2007).

15. Harcombe, E. S. & Wyman, R. J. Output pattern generation by *Drosophila* flight motoneurons. *J. Neurophysiol.* **40**, 1066–1077 (1977).

16. Ryglewski, S. et al. Dendrites are dispensable for basic motoneuron function but essential for fine tuning of behavior. *Proc. Natl Acad. Sci. USA* **111**, 18049–18054 (2014).

17. Wang, Q., Zhao, C. & Swank, D. M. Calcium and stretch activation modulate power generation in *Drosophila* flight muscle. *Biophys. J.* **101**, 2207–2213 (2011).

18. Harcombe, E. S. & Wyman, R. J. The cyclically repetitive firing sequences of identified *Drosophila* flight motoneurons. *J. Comp.*

Physiol. **123**, 271–279 (1978).

19. Selverston, A. I. Invertebrate central pattern generator circuits. *Philos. Trans. R. Soc. B* **365**, 2329–2345 (2010).

20. Kiehn, O. Decoding the organization of spinal circuits that control locomotion. *Nat. Rev. Neurosci.* **17**, 224–238 (2016).

21. Ryglewski, S., Vonhoff, F., Scheckel, K. & Duch, C. Intra-neuronal competition for synaptic partners conserves the amount of dendritic building material. *Neuron* **93**, 632–645 (2017).

22. Koenig, J. H. & Ikeda, K. Neural interactions controlling timing of flight muscle activity in *Drosophila*. *J. Exp. Biol.* **87**, 121–136 (1980).

23. Koenig, J. H. & Ikeda, K. Reciprocal excitation between identified flight motor neurons in *Drosophila* and its effect on pattern generation. *J. Comp. Physiol. A* **150**, 305–317 (1983).

24. Coombes, S. Neuronal networks with gap junctions: a study of piecewise linear planar neuron models. *SIAM J. Appl. Dyn. Syst.* **7**, 1101–1129 (2008).

25. Phelan, P. Innexins: members of an evolutionarily conserved family of gap-junction proteins. *Biochim. Biophys. Acta* **1711**, 225–245 (2005).

26. Goodenough, D. A. & Paul, D. L. Gap junctions. *Cold Spring Harb. Perspect. Biol.* **1**, a002576 (2009).

27. Connors, B. W. Synchrony and so much more: diverse roles for electrical synapses in neural circuits. *Dev. Neurobiol.* **77**, 610–624 (2017).

28. Kennedy, T. & Broadie, K. Fragile X mental retardation protein restricts small dye iontophoresis entry into central neurons. *J. Neurosci.* **37**, 9844–9858 (2017).

29. Alcamí, P. & Pereda, A. E. Beyond plasticity: the dynamic impact of electrical synapses on neural circuits. *Nat. Rev. Neurosci.* **20**, 253–271

(2019).

30. Chow, C. C. & Kopell, N. Dynamics of spiking neurons with electrical coupling. *Neural Comput.* **12**, 1643–1678 (2000).

31. Lewis, T. J. & Rinzel, J. Dynamics of spiking neurons connected by both inhibitory and electrical coupling. *J. Comput. Neurosci.* **14**, 283–309 (2003).

32. Ostojic, S., Brunel, N. & Hakim, V. Synchronization properties of networks of electrically coupled neurons in the presence of noise and heterogeneities. *J. Comput. Neurosci.* **26**, 369–392 (2009).

33. Han, S. K., Kurrer, C. & Kuramoto, Y. Dephasing and bursting in coupled neural oscillators. *Phys. Rev. Lett.* **75**, 3190–3193 (1995).

34. Vervaeke, K. et al. Rapid desynchronization of an electrically coupled interneuron network with sparse excitatory synaptic input. *Neuron* **67**, 435–451 (2010).

35. Zanette, D. H. Synchronization and frustration in oscillator networks with attractive and repulsive interactions. *Europhys. Lett.* **72**, 190–196 (2005).

36. Zillmer, R., Livi, R., Politi, A. & Torcini, A. Stability of the splay state in pulse-coupled networks. *Phys. Rev. E* **76**, 046102 (2007).

37. Hesse, J., Schleimer, J.-H. & Schreiber, S. Qualitative changes in phase-response curve and synchronization at the saddle-node-loop bifurcation. *Phys. Rev. E* **95**, 052203 (2017).

38. Hesse, J., Schleimer, J.-H., Maier, N., Schmitz, D. & Schreiber, S. Temperature elevations can induce switches to homoclinic action potentials that alter neural encoding and synchronization. *Nat. Commun.* **13**, 3934 (2022).

39. Berger, S. D. & Crook S. M. Modeling the influence of ion channels on neuron dynamics in *Drosophila*. *Front. Comput. Neurosci.* **9**, 139

(2015).

40. Izhikevich, E. M. *Dynamical Systems in Neuroscience* (MIT Press, 2007).

41. Marder, E., Gutierrez, G. J. & Nusbaum, M. P. Complicating connectomes: electrical coupling creates parallel pathways and degenerate circuit mechanisms. *Dev. Neurobiol.* **77**, 597–609 (2017).

42. Marder, E. Roles for electrical coupling in neural circuits as revealed by selective neuronal deletions. *J. Exp. Biol.* **112**, 147–167 (1984).

43. Weaver, A. L., Roffman, R. C., Norris, B. J. & Calabrese, R. L. A role for compromise: synaptic inhibition and electrical coupling interact to control phasing in the leech heartbeat CPG. *Front. Behav. Neurosci.* **4**, 38 (2010).

44. Gutierrez, G. J., O'Leary, T. & Marder, E. Multiple mechanisms switch an electrically coupled, synaptically inhibited neuron between competing rhythmic oscillators. *Neuron* **77**, 845–858 (2013).

45. Johnson, B. R., Peck, J. H. & Harris-Warrick, R. M. Amine modulation of electrical coupling in the pyloric network of the lobster stomatogastric ganglion. *J. Comp. Physiol. A* **172**, 715–732 (1993).

46. Sevetson, J. & Haas, J. S. Asymmetry and modulation of spike timing in electrically coupled neurons. *J. Neurophysiol.* **113**, 1743–1751 (2015).

47. Jenett, A. et al. A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991–1001 (2012).

48. Dawydow, A. et al. Channelrhodopsin-2-XXL, a powerful optogenetic tool for low-light applications. *Proc. Natl Acad. Sci.* **111**, 13972–13977 (2014).

49. Zhang, Y. et al. Fast and sensitive GCaMP calcium indicators for imaging neural populations. *Nature* **615**, 884–891 (2023).

50. Barthmaier, P. & Fyrberg, E. Monitoring development and pathology of *Drosophila* indirect flight muscles using green fluorescent protein. *Dev. Biol.* **169**, 770–774 (1995).

51. Islas, L. D. & Sigworth, F. J. Voltage sensitivity and gating charge in Shaker and Shab family potassium channels. *J. Gen. Physiol.* **114**, 723–742 (1999).

52. Ryglewski, S. & Duch, C. Preparation of *Drosophila* central neurons for in situ patch clamping. *J. Vis. Exp.* https://doi.org/10.3791/4264 (2012).

53. Stimberg, M., Brette, R. & Goodman, D. F. Brian 2, an intuitive and efficient neural simulator. *eLife* **8**, e47314 (2019).

54. Doedel, E. J. et al. AUTO-07P: continuation and bifurcation software for ordinary differential equations. https://www.macs.hw.ac.uk/~gabriel/auto07/auto.html (2007).

55. Niemeyer, N., Schleimer, J.-H. & Schreiber, S. Biophysical models of intrinsic homeostasis: firing rates and beyond. *Curr. Opin. Neurobiol.* **70**, 81–88 (2021).

56. Kirst, C. *Synchronization, Neuronal Excitability and Information Flow in Networks of Neuronal Oscillators*. (Niedersächsische Staats and Universitätsbibliothek Göttingen, 2012).

57. Brown, E., Moehlis, J. & Holmes, P. On the phase reduction and response dynamics of neural oscillator populations. *Neural Comput.* **16**, 673–715 (2004).

58. Ermentrout, B. & Type, I. Membranes, phase resetting curves, and synchrony. *Neural Comput.* **8**, 979–1001 (1996).

59. Schleimer, J.-H. & Schreiber, S. Phase-response curves of ion channel gating kinetics. *Math. Methods Appl. Sci.* https://doi.org/10.1002/mma.5232 (2018).

60. Ronge, R. & Zaks, M. A. Splay states and two-cluster states in ensembles of excitable units. *Eur. Phys. J. Spec. Top.* **230**, 2717–2724 (2021).

61. Ermentrout, G. B. & Kopell, N. Parabolic bursting in an excitable system coupled with a slow oscillation. *SIAM J. Appl. Math.* **46**, 233–253 (1986).

62. Galán, R. F., Ermentrout, G. B. & Urban, N. N. Efficient estimation of phase-resetting curves in real neurons and its significance for neural-network modeling. *Phys. Rev. Lett.* **94**, 158101 (2005).

63. Hesse, J. & Schreiber, S. How to correctly quantify neuronal phase-response curves from noisy recordings. *J. Comput. Neurosci.* **47**, 17–30 (2019).

64. Kuehn, C. & Duch, C. Putative excitatory and putative inhibitory inputs are localised in different dendritic domains in a *Drosophila* flight motoneuron. *Eur. J. Neurosci.* **6**, 860–875 (2013).

65. Talay, M. et al. Transsynaptic mapping of second-order taste neurons in flies by trans-tango. *Neuron* **96**, 783–795 (2017).

Acknowledgements

This project was supported by the DFG Research Unit 5289 (DU 331/15-1, RY 117/4-1, SCHR 1239/5-1); additional DFG funding to C.D. (DU 331/6-3) and to S.R. (RY 117/3-2); funding by the Carl-Zeiss-Foundation to S.H.; the Einstein Foundation Berlin (EP-2021-621) to S.S.; and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 864243) to S.S. We thank O. Buda and A. Buijnenko for help with tethered flight recordings; U. Thomas for help with the cloning of UAS-Shab; P. R. Hiesinger, U. Thomas and M. Silies for comments on the manuscript; and F. Sigworth for providing the *Shab* cDNA.

Author information

Author notes

1. These authors contributed equally: Silvan Hürkey, Nelson Niemeyer

2. These authors jointly supervised this work: Susanne Schreiber, Carsten Duch

Authors and Affiliations

1. Institute of Developmental Biology and Neurobiology (iDN), Johannes Gutenberg-University Mainz, Mainz, Germany

   Silvan Hürkey, Stefanie Ryglewski & Carsten Duch

2. Institute for Theoretical Biology, Humboldt-Universität zu Berlin, Berlin, Germany

   Nelson Niemeyer, Jan-Hendrik Schleimer & Susanne Schreiber

3. Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany

   Susanne Schreiber

Contributions

S.H. conducted all in vivo recordings and all calcium imaging recordings. N.N. and J.-H.S. conducted and analysed the mathematical modelling. S.H. and N.N. analysed the in vivo recordings. S.R. conducted and analysed all patch-clamp data. Immunohistochemical experiments and CLSM were conducted and analysed by S.H. and C.D. All of the authors helped to conceptualize the work. C.D. and S.R. supervised the experimental work, J.-H.S. and S.S. supervised the theoretical work. C.D. and S.S. wrote the manuscript. All of the authors helped to edit the text and the figures.

Corresponding authors

Correspondence to Susanne Schreiber or Carsten Duch.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks David Schulz, John Tuthill and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Firing phase relationships of MN1–5 pairs are conserved between individuals.

For three individual male flies **(a–c)** all 10 possible pairwise combinations of the DLM-MNs (MN1–5) are plotted as cumulative phase histograms from simultaneous recordings of MN1–5 during 10 min of continuous tethered flight (equalling approximately 3000 spikes of each MN). Starting with the MN1/MN2 pair on the upper left, **(a1-c1)**, for each MN pair, the interspike intervals between two consecutive spikes of one MN were divided into 100 equally sized bins (x-axis), and it was determined in which bin the spike of the other MN occurred. This was repeated as sliding window for all interspike intervals to fill the bins cumulatively. Cumulative spike counts in each bin were normalized to total spike count (y-axis). In all three individuals **(a–c)** different MN pairs show different phase relationships, but the same pairs show similar phase relationships across

individuals. The phase histograms of all MN pairs show a characteristic gap around phases zero and 1, indicating preferred out-of-phase firing. Apart from this characteristic gap, phase relations are not highly precise, so that some distributions are broad. This is owed to multiple preferred MN1–5 firing sequences (splay states) occurring within one flight bout (see Fig. 1 in the main text). **(d)** compares phase histograms restricted to one splay state and with changing splay states (both from the same animal and 1 min of flight). Histograms restricted to one splay state show a much narrower phase relation and no synchronous spikes at phases 0 or 1.

Extended Data Fig. 2 Phase relations between MN1–5 are conserved across species.

In all tested flying insect species, the dorsal longitudinal wing depressor muscles (DLM) consists of 6 muscle fibres, each of which is innervated by one MN, namely MN1–5 (for schematic see Fig. 1a). In *Drosophila melanogaster* each of the 10 MN1–5 pairs shows characteristic phase relationships in their firing patterns (see Fig. 1 in the main text, Extended Data Fig. 1). Analyses of *in vivo* recordings during tethered flight reveals highly similar firing phase relationships of different MN pairs in *Drosophila melanogaster* **(a)** and in the goldfly, *Lucilia spec* **(b)**, a related dipteran species. **(c)** *In vivo* recordings of the MN4-MN5 pair shows similar firing phase relationships between the fruit fly *Drosophila melanogaster* (top), another Drosophilidae (*Drosophila hydei*, second from top), two additional dipteran species (the blowfly *Calliphora spec*., third from top; and the house fly, *Musca domestica*, second from bottom). Some characteristics of the phase relationships observed in all dipteran species tested (top 4 panels) are also observed in a hymenopteran species, the honey bee *Apis mellifera* (bottom). There the characteristic inhibition of MN4 firing just after MN5 spikes is recapitulated, but the depression of MN4 firing just before MN5 spikes is not.

Extended Data Fig. 3 Unpatterned excitatory input scales up CPG output without affecting the pattern.

(a) Representative recordings of MN4 (arrow, smaller unit) and MN5 (arrow, large unit) with one tungsten electrode during tethered flight and blue light stimulation in a fly with expression of UAS-Channelrhodopsin in cholinergic neurons (Chat-GAL4xUAS-XXL-ChR, second from bottom; 11 replicate animals) and in a control fly (Chat-GAL4x w^{1118}, bottom trace; 3 replicate animals). Blue bars indicate the timing of blue light stimulation. Upper three traces show the instantaneous wingbeat frequency (top trace) and the instantaneous firing frequencies of MN5 and MN4 upon optogenetic activation of presynaptic cholinergic neurons. **(ai, left)** Quantification from 11 animals with expression of UAS-ChR in cholinergic neurons reveals that the firing frequencies of MN4 and MN5 are increased significantly and to the same degree during optogenetic stimulation (blue circles) as compared to before light stimulation (green circles). By contrast, in controls (Chat-GAL4 x w1118; n = 3) light stimulation has no effect. **(ai, right)** Average firing frequency increases in MN4 and MN5 (black dots) but not in controls (grey dots, horizontal bars depict means, error bars the SDs). **(aii)** Consequently, wingbeat frequency is significantly increased upon optogenetic activation of cholinergic neurons (11 animals, horizontal bar depicts mean, error bars the SD), but not upon light stimulation in controls (3 animals). **(b)** The typical phase relationship of MN4 and MN5 firing that is observed in controls (Fig. 2a,b in the main text, Extended Data Fig. 1) and prior to optogenetic stimulation (left) remains unaltered upon increasing CPG output by unpatterned stimulation of presynaptic cholinergic interneurons (right). Coloured bars represent the average values from 11 animals and grey bars the SD.

Extended Data Fig. 4 MN1–5 receive common excitatory cholinergic input and control wingbeat rate.

(a) Representative confocal projection view (5 replicate animals) of the VNC flight motor neuropil in an animal with expression of UAS-CD4 td tomato (magenta) and one half of split-GFP (UAS-CD4-spGFP11) under the control of GAL4 in DLM-MNs (driven with GMRH2306-ADZ attP40; GMR30A07-DBD attP2, see also Extended Data Fig. 5a,b for expression patterns with UAS-GFP). The other half of split-GFP (spGFP1-10) is coupled to the synaptic vesicle protein synaptobrevin (nSyb) and driven under the control of LexA in cholinergic neurons (ChaT-LexA x lexAop-

nSyb-spGFP1-10). GFP reconstitution across synaptic partners (GRASP) occurs upon release of synaptic vesicles from cholinergic neurons and results in GFP fluorescence at the sites where nSyb- spGFP1-10 is released in direct proximity of MN1–5 dendritic membrane with expression of CD4-spGFP11 (green). **(b)** Selective enlargement of synaptic activity dependent GRASP signal (green) at contacts between ChaT positive cholinergic neurons and td tomato expressing MN dendrites (magenta) indicates direct synaptic inputs from cholinergic neurons to MN1–5. **(c)** Dual current clamp recording from 2 MNs (here MN3 and MN4) show highly similar patterns of PSPs, providing direct evidence for shared synaptic input (representative traces from 1 animal, 4 replicate animals). **(d)** Representative example for similar changes in firing frequency of MN1–5 indicates shared synaptic drive (8 replicate animals). **(e)** Wingbeat frequencies and MN1–5 firing frequencies are calculated for 1-second bins within flight. Flight starts and stops are omitted to avoid trivial correlations. The correlation between MN1–5 firing frequency and wingbeat frequency is moderate across animals, though significant (two-sided Pearson correlation coefficient $r = 0.31$, $p < 0.0001$). **(f)** In the same animal, changes in MN1–5 (ΔMN) and wingbeat frequency (Δwb, comparing consecutive 1-second bins) show a stronger correlation (Pearson $r = 0.6$, $p < 0.0001$). **(g)** The ΔMN/Δwb correlation is also stronger across animals.

Extended Data Fig. 5 Unpatterned optogenetic activation of MN1–5 during flight scales up CPG output without affecting the pattern.

(a) Representative projection views of confocal image stacks (5 replicate animals) as overview of the entire ventral nerve cord (VNC, outlined by dotted line) and **(b)** as selective enlargement of the pterothoracic VNC with a expression of UAS-CD4-td::GFP under the control of DLM MN Split-GAL4 (GMR23H06-ADZ attP49; GMR30A07-DBD attP2). The VNC midline is indicated by a dotted white line, MN1-4 somata are labelled by white arrows on both sides of the VNC. MN5 somata (white arrows) are located closer to the midline and partially obscured by MN1–5 dendrites in the flight motor neuropil. The only other signals detected are due to autofluorescence from cuticular structures. **(c,d)** Optogenetic activation of MN1–5 during flight. **(c)** The bottom trace shows a representative recording

of MN4 (arrow, smaller unit) and MN5 (arrow, large unit) with one tungsten electrode during tethered flight. Timing of unpatterned optogenetic activation of MN1–5 is indicated by a blue line. Upper three traces show the instantaneous wingbeat frequency (top trace) and the instantaneous firing frequencies of MN5 and MN4 (7 replicate animals). **(d)** Quantification shows that optogenetic activation increases the firing frequencies of both MNs significantly and to the same degree (horizontal bar depicts mean, error bars the SD). **(di)** Consequently, wingbeat frequency is increased significantly (n = 7 animals, horizontal bar depicts mean, error bars the SD). **(e)** The typical phase relationship of MN4 and MN5 firing that is observed in controls (Fig. 2a, Extended Data Fig. 1) and prior to optogenetic stimulation (left) remains unaltered upon increasing CPG output by unpatterned stimulation the MN1–5 (right). Coloured bars represent the average values from 7 animals and grey bars the s.e.m.

Extended Data Fig. 6 Dip around zero in phase histograms is an appropriate indicator for splayness and not affected by inhibitory chemical synapses.

(**a**) Phase histograms for simulated MN1–5 with homogeneous coupling (occ [%] is the percentage of events in each phase bin). All-to-all rather than single-pair histograms are chosen to mitigate initial condition effects for finite run times. For low noise 4 peaks correspond to the 4 phase relations in a perfect splay state. With increasing noise, the dip around zero is more persistent than the subsequent dips. **(b)** Similar all-to-all phase histogram for an example experimental control. **(c)** Washing out of the dip around zero corresponds to decreased network splayness (y-axis, splayness index). For comparison, mean and standard deviation of experimentally measured control and Canton-S flies are shown (n = 8 animals). **(d)** For MN4-5 histograms the dip around zero is quantified using the fraction that falls within (−0.1,0.1). This fraction is negatively correlated to full network splayness in both simulated (considering all simulations from Fig. 3d,h) and experimental data where all 5 neurons were available (Canton-S, experimental control (GMR23H06-ADZ attP49;GMR30A07-DBD attP2 > P{UAS-GFP.VALIUM10}attP2), ShakB RNAi-kd (GMR23H06-ADZ attP49;GMR30A07-DBD attP2 > P{TRiP.HMC04895}attP2)). For the experimental data, a two-sided Pearson correlation determined a strong

negative correlation ($r = -0.94$, $p < 0.0001$, $n = 10$). **(e,f)** Firing phase relationships between MN4 and MN5 in control animals (see also Fig. 2a,b, Extended Data Fig. 1) remain similar upon UAS-RNAi knockdown of receptors for inhibitory chemical synapses in MN1–5 (under the control of DLM-MN spilt-GAL4, *GMR23H06-ADZ attP49; GMR30A07-DBD attP2*), **(e)** the glutamate gated chloride channel (GluCl) and **(f)** Rdl GABA-ARs. GluCl-RNAi knockdown efficacy was confirmed by Western blotting and Rdl GABA-AR knockdown efficacy has previously been confirmed[64]. Coloured bars represent the average values from 10 animals for GluCl-RNAi, 8 animals for Rdl GABA-AR-RNAi, and grey bars the s.e.m.

Extended Data Fig. 7 No detection of output synapses from MN1–5 in the ventral nerve cord (VNC).

(a) Representative (from 4 replicate animals) trans-synaptic labelling of postsynaptic partners (**a, aii** magenta) of LC visual interneurons (**a, ai**, green) in the adult fly brain with Trans-Tango[65]. **(b)** Representative (from 4 replicate animals) trans-synaptic labelling of postsynaptic partners (**b, bii** magenta) of period interneurons in the adult VNC (**b, bi**, green). **(c)** By contrast trans-synaptic labelling with Trans-Tango reveals no postsynaptic partners (**c, cii**, magenta) of MN1–5 (**c, ci**, green) in the adult fly VNC (6 replicate animals). **(d)** Expression of Brp-short (w;23H06-ADZ UAS-Brp.S^{straw-}D3/+;30A07-DBD/+;) tagged with strawberry in MN1–5 (4 replicate animals) causes Brp-short strawberry puncta in the MN1–5 somata, where the protein is produced, but not in any central arbours (**d, di, dii**, MN1–5 green, Brp short magenta). By contrast, in the same preparation, Brp-short puncta (white arrows) localize to presynaptic active zones at MN axon terminals at the neuromuscular junction (**e, ei, eii**, MN1–5 green, Brp short magenta).

Extended Data Fig. 8 MN1–5 firing phase relationships require gap junctions (GJs) between MNs.

For three individual male *Drosophila melanogaster* flies with targeted expression of RNAi knockdown for ShakB in MN1–5 **(a–c)** all 10 possible pairwise combinations of the MN1–5 are plotted as cumulative phase

histograms from simultaneous recordings of MN1–5 during 10 min of continuous tethered flight (equalling approximately 3000 spikes of each MN). Starting with the MN1-MN2 pair on the upper left, **(a1-c1)**, for each MN pair, the interspike intervals between two consecutive MN spikes was normalized and divided into 100 equally sized bins (x-axis). Next it was determined in which bin the spike of the other MN occurred. This was repeated as sliding window for all interspike intervals and bins were filled cumulatively and normalized to total spike count (y-axis). In all three individuals with knockdown for ShakB mediated GJs in MN1–5 **(a–c)** the phase relationships that are characteristic for each MN pair in control flies (see Extended Data Fig. 1) and conserved across dipteran species (see Extended Data Fig. 2) are impaired.

Extended Data Fig. 9 Motoneuron model analysis and phase-reduced view on coupling.

(a) Codim-2 bifurcation diagram of the motoneuron model with focus on global bifurcations that switch the PRC shape with changing Shab levels (y-axis, g_{shab}). Small and big saddle-node loop (sSNS, bSNL) encapsulate the SNIC interval, where spiking (grey area) commences after a saddle-node bifurcation (blue line) on a limit cycle. Saddle Homoclinic loop (HOM) line is green. The Andronov-Hopf bifurcation (AH, red line), ends in a Bogdanov-Takens (BT) point. At the cusp (CP) the neuron model switches from 3 to 1 fixpoints. The three PRCs (insets, right) show distinct symmetry properties. **(b)** Simplified illustration how asymmetric PRC components induce stable phase relations. With spike shapes in the form of delta functions, the odd parts of the PRC and of the coupling function have the same shape. The mean PRC component is absorbed into the average firing rate and plotted is only the PRC’s odd part (grey). Upon perturbations (arrows) PRCs with phase-advance in the ISI’s first half and phase-delay in the second half cause anti-phase synchronization (top), while the reverse causes in-phase synchronization (bottom). **(c)** Iterated maps (cobweb plots) illustrate the qualitative behaviour of phase differences. Starting at a random phase difference, ψ_n, a vertical arrow is drawn from the diagonal to the function curve, constituting the new phase, ψ_{n+1}, after a perturbation. From this point a horizontal arrow is drawn back to the diagonal, such that

ψ_n is substituted by ψ_{n+1}. Repeating these two steps, the system converges to stable fixpoints in phase differences (according to equation (3); Methods). In the case of a PRC with phase-advance in the ISI's first half and phase-delay in the second half (left), the fix point is $\psi = 0.5$, while for the reverse case (right) the fix point is $\psi = 0$.

Extended Data Fig. 10 Shab constitutes 50% of the delayed rectifier current in MN1–5.

(a) Outward current as measured in control MNs in TTX containing (10^{-7} M) calcium free saline by applying command voltage steps from a holding potential of −20 mV down to −60 mV and up to +50 mV in 10 mV increments (top traces). Same recording in 100 µM quinidine (middle). Bottom traces show the quinidine-sensitive Shab current by subtraction of the top minus the middle traces. **(b)** Same experiment but following Shab RNAi-kd (Shab-kd) in MN1–5. **(c)** Same experiment but following UAS-Shab overexpression (Shab-oe) in MN1–5. **(d)** Quantification of total delayed rectifier (DR) outward current (top), quinidine insensitive, non-Shab DR current (middle), and Shab DR current (bottom) in control (black, 6 replicate animals), with Shab-RNAi (blue, 11 replicate animals), and following Shab-overexpression in MN1–5 (red, 7 replicate animals, 5 for isolated Shab and non Shab current). In control, Shab constitutes ~50% of the DR current. Shab-RNAi significantly reduces Shab current by ~70% (ANOVA with Tukey's posthoc comparison, $p = 0.0016$), but does not affect quinidine insensitive non Shab current (ANOVA with Tukey's posthoc comparison, $p = 0.8215$) or total DR current (ANOVA with Tukey's posthoc comparison, $p = 0.0608$). Shab overexpression significantly increases Shab current amplitude (ANOVA with Tukey's posthoc comparison, $p = 0.0029$), significantly increases total DR current (ANOVA with Tukey's posthoc comparison, $p = 0.0010$), but non Shab DR current remains unaffected (ANOVA with Tukey's posthoc comparison, $p = 0.6291$).

Supplementary information

Supplementary Information

Supplementary Tables 1–5, listing reagents for experiments, parameters for computational modelling and the genotypes of all of the fly strains used.

Reporting Summary

Supplementary Video 1

Video *of Drosophila melanogaster* during tethered flight. High-speed video (5,000 fps) of *D. melanogaster* during tethered flight with one recording electrode.

Supplementary Video 2

Video of *Lucilia* sp. during tethered flight. High-speed video (5,000 fps) of *Lucilia* sp. during tethered flight with 2 recording electrodes.

Supplementary Video 3

Video of *A. mellifera* during tethered flight. High-speed video (5,000 fps) of *A. mellifera* during tethered flight.

Supplementary Video 4

In vivo calcium imaging from the DLM fibres of *D. melanogaster.* In vivo calcium imaging from DLM fibre 6 with the expression of GCaMP8f (*w[*];P{y[+t7.7] w[+mC]=20XUAS-IVS-jGCaMP8f}su(Hw)attP5/+;P{[+mC]=Act88F-GAL4.1.3}3/+*) in response to electrical stimulation with different frequencies (0.5, 1, 2, 5 and 10 Hz). The field of view contains the dorsal thorax with the scutellum (purple line) at the posterior end and the head anteriorly (out of focus). The DLM fibre on one side is encircled (cyan) and the region of interest for $\Delta F/F$ GCaMP8f intensity measurements is indicated by a green circle. Calcium responses of the stimulated DLM fibre are shown as a video in real time, and $\Delta F/F$ values are played synchronously on the top.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Hürkey, S., Niemeyer, N., Schleimer, JH. *et al.* Gap junctions desynchronize a neural circuit to stabilize insect flight. *Nature* **618**, 118–125 (2023). https://doi.org/10.1038/s41586-023-06099-0

- Received: 18 February 2022
- Accepted: 18 April 2023
- Published: 24 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06099-0

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

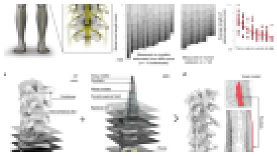
Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

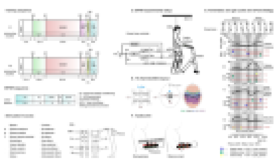
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06099-0



Activity-dependent spinal cord neuromodulation rapidly restores trunk and leg motor functions after complete paralysis

07 February 2022

Andreas Rowald, Salif Komi, … Grégoire Courtine



Non-invasive, Brain-controlled Functional Electrical Stimulation for Locomotion Rehabilitation in Individuals with Paraplegia

01 May 2019

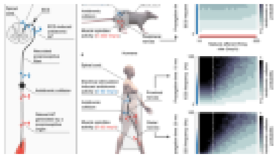
Aurelie Selfslagh, Solaiman Shokur, … Miguel A. L. Nicolelis



Targeted neurotechnology restores walking in humans with spinal cord injury

31 October 2018

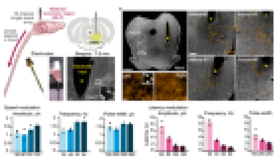
Fabien B. Wagner, Jean-Baptiste Mignardot, … Grégoire Courtine



[Electrical spinal cord stimulation must preserve proprioception to enable locomotion in humans with spinal cord injury]

31 October 2018

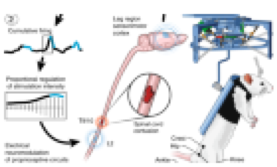
Emanuele Formento, Karen Minassian, … Gregoire Courtine



[Multi-pronged neuromodulation intervention engages the residual motor circuitry to facilitate walking in a rat model of spinal cord injury]

26 March 2021

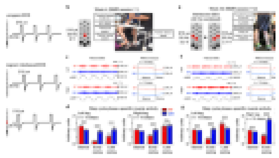
Marco Bonizzato, Nicholas D. James, … Gregoire Courtine



[Brain-controlled modulation of spinal circuits improves recovery from spinal cord injury]

01 August 2018

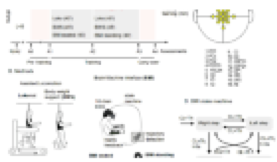
Marco Bonizzato, Galyna Pidpruzhnykova, … Grégoire Courtine



[Neuromodulation of lumbosacral spinal networks enables independent stepping after complete paraplegia]

24 September 2018

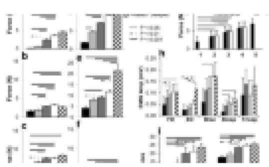
Megan L. Gill, Peter J. Grahn, … Kristin D. Zhao



[Training with noninvasive brain–machine interface, tactile feedback, and locomotion to enhance neurological recovery in individuals with complete paraplegia: a randomized pilot study]

29 November 2022

Miguel A. L. Nicolelis, Eduardo J. L. Alho, … Solaiman Shokur



[Engaging cervical spinal circuitry with non-invasive spinal stimulation and buspirone to restore hand function in chronic motor complete patients]

19 October 2018

Yevgeniy Freyvert, Nicholas Au Yong, … Daniel C. Lu

- Article

- Open Access
- Published: 24 May 2023

Walking naturally after spinal cord injury using a brain–spine interface

- Henri Lorach,
- Andrea Galvez,
- Valeria Spagnolo,
- Felix Martel,
- Serpil Karakas,
- Nadine Intering,
- Molywan Vat,
- Olivier Faivre,
- Cathal Harte,
- Salif Komi,
- Jimmy Ravier,
- Thibault Collin,
- Laure Coquoz,
- Icare Sakr,
- Edeny Baaklini,
- Sergio Daniel Hernandez-Charpak,
- Gregory Dumont,
- Rik Buschman,
- Nicholas Buse,
- Tim Denison,
- Ilse van Nes,
- Leonie Asboth,
- Anne Watrin,
- Lucas Struber,
- Fabien Sauter-Starace,
- Lilia Langar,
- Vincent Auboiroux,
- Stefano Carda,
- Stephan Chabardes,

- Tetiana Aksenova,
- Robin Demesmaeker,
- Guillaume Charvet ,
- Jocelyne Bloch &
- …
- Grégoire Courtine

Show authors

Nature **volume 618**, pages 126–133 (2023)

- 231k Accesses
- 1 Citations
- 6299 Altmetric
- Metrics details

Abstract

A spinal cord injury interrupts the communication between the brain and the region of the spinal cord that produces walking, leading to paralysis[1,2]. Here, we restored this communication with a digital bridge between the brain and spinal cord that enabled an individual with chronic tetraplegia to stand and walk naturally in community settings. This brain–spine interface (BSI) consists of fully implanted recording and stimulation systems that establish a direct link between cortical signals[3] and the analogue modulation of epidural electrical stimulation targeting the spinal cord regions involved in the production of walking[4,5,6]. A highly reliable BSI is calibrated within a few minutes. This reliability has remained stable over one year, including during independent use at home. The participant reports that the BSI enables natural control over the movements of his legs to stand, walk, climb stairs and even traverse complex terrains. Moreover, neurorehabilitation supported by the BSI improved neurological recovery. The participant regained the ability to walk with crutches overground even when the BSI was switched off. This

digital bridge establishes a framework to restore natural control of movement after paralysis.

Main

To walk, the brain delivers executive commands to the neurons located in the lumbosacral spinal cord[7]. Although the majority of spinal cord injuries do not directly damage these neurons, the disruption of descending pathways interrupts the brain-derived commands that are necessary for these neurons to produce walking[8]. The consequence is permanent paralysis.

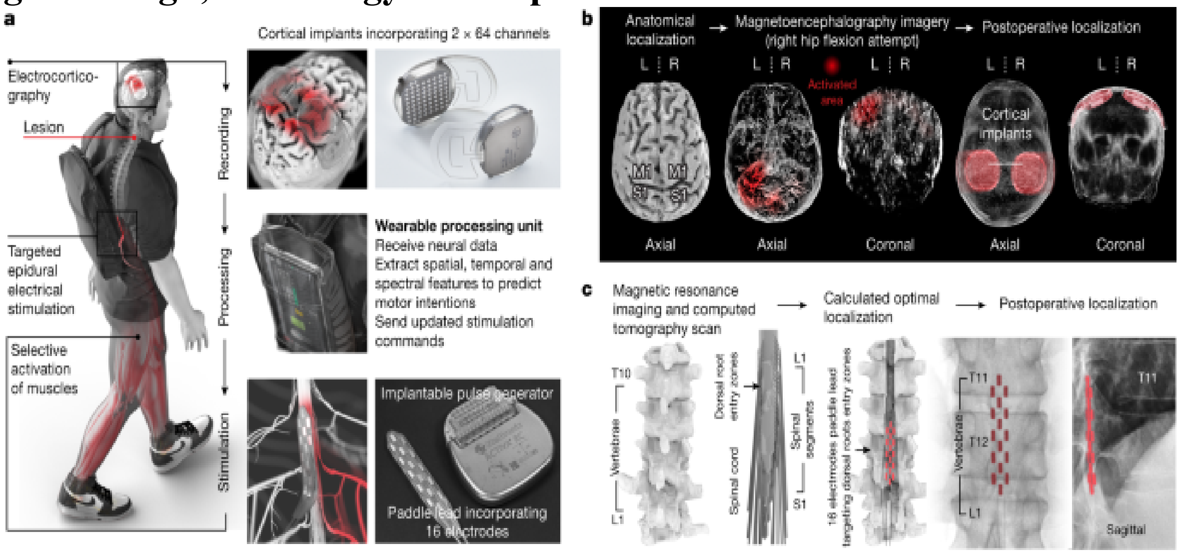
We previously showed that epidural electrical stimulation targeting the individual dorsal root entry zones of the lumbosacral spinal cord enables the modulation of specific leg motor pools[9,10,11,12]. In turn, recruiting these dorsal root entry zones with preprogrammed spatiotemporal sequences replicates the physiological activation of leg motor pools underlying standing and walking[4,5,11,13,14]. These stimulation sequences restored standing and basic walking in people with paralysis due to a spinal cord injury. However, this recovery required wearable motion sensors to detect motor intentions from residual movements or compensatory strategies to initiate the preprogrammed stimulation sequences[5]. Consequently, the control of walking was not perceived as completely natural. Moreover, the participants showed limited ability to adapt leg movements to changing terrain and volitional demands.

Here, we suggest that a digital bridge[13,15,16,17,18,19] between the brain and spinal cord would enable volitional control over the timing and amplitude of muscle activity, restoring more natural and adaptive control of standing and walking in people with paralysis due to spinal cord injury.

Digital bridge from brain to spinal cord

To establish this digital bridge, we integrated two fully implanted systems that enable recording of cortical activity and stimulation of the lumbosacral spinal cord wirelessly and in real time (Fig. 1a).

Fig. 1: Design, technology and implantation of the BSI.



a, Two cortical implants composed of 64 electrodes are positioned epidurally over the sensorimotor cortex to collect ECoG signals. A processing unit predicts motor intentions and translates these predictions into the modulation of epidural electrical stimulation programs targeting the dorsal root entry zones of the lumbosacral spinal cord. Stimulations are delivered by an implantable pulse generator connected to a 16-electrode paddle lead. **b**, Images reporting the pre-operative planning of cortical implant locations, and postoperative confirmation. L, left; R, right. **c**, Personalized computational model predicting the optimal localization of the paddle lead to target the dorsal root entry zones associated with lower limb muscles, and postoperative confirmation.

To monitor electrocorticographic (ECoG) signals from the sensorimotor cortex, we leveraged the WIMAGINE technology[3,20]. WIMAGINE implants consist of an 8-by-8 grid of 64 electrodes (4 mm × 4.5 mm pitch in anteroposterior and mediolateral axes, respectively) and recording electronics that are embedded within a 50 mm diameter, circular-shaped titanium case that has the same thickness as the skull. The geometry of the system favours close and stable contact between the electrodes and the dura mater, and renders the devices invisible once implanted within the skull.

Two external antennas are embedded within a personalized headset that ensures reliable coupling with the implants. The first antenna powers the implanted electronics through inductive coupling (high frequency, 13.56

MHz), whereas the second, ultrahigh frequency antenna (UHF, 402–405 MHz) transfers ECoG signals in real time to a portable base station and processing unit, which generates online predictions of motor intentions on the basis of these signals (Extended Data Fig. 1).

The decoded motor intentions are then converted into stimulation commands that are transferred to tailored software running on the same processing unit.

These commands are delivered to the ACTIVA RC implantable pulse generator (Fig. 1a), which is commonly used to deliver deep brain stimulation in patients with Parkinson’s disease. We upgraded this implant with wireless communication modules that enabled real-time adjustment over the location and timing of epidural electrical stimulation with a latency of about 100 ms (Extended Data Fig. 1).

Electrical currents are then delivered to the targeted dorsal root entry zones using the Specify 5-6-5 implantable paddle lead, which consists of an array incorporating 16 electrodes.

This integrated chain of hardware and software established a wireless digital bridge between the brain and the spinal cord: a brain–spine interface (BSI) that converts cortical activity into the analogue modulation of epidural electrical stimulation programs to tune lower limb muscle activation, and thus regain standing and walking after paralysis due to a spinal cord injury (Supplementary Video 1).

Neurosurgical implantation of the BSI

In the context of the Stimulation Movement Overground (STIMO)-BSI clinical trial (clinicaltrials.gov, NCT04632290), we enroled a 38-year-old male who had sustained an incomplete cervical (C5/C6) spinal cord injury during a biking accident ten years before. He had previously participated in the STIMO clinical trial (clinicaltrials.gov, NCT02936453), which involved a five-month neurorehabilitation programme supported by targeted epidural electrical stimulation of the spinal cord[4,5]. This programme enabled him to regain the ability to step with the help of a front-wheel walker. Despite continued use of the stimulation at home, for approximately three years, he

had reached a neurological recovery plateau, which motivated him to enrol in STIMO-BSI.

To guide the implantation of the BSI, we developed pre-operative planning procedures that enabled us to optimize the positioning of the recording and stimulation implants over the brain and spinal cord.

The BSI requires the detection of neural features related to the intention to move the left and right lower limbs. To identify the cortical regions most responsive to the attempt to move each joint of the lower limbs, we acquired anatomical and functional imaging data based on computerized tomography and magnetoencephalography (Fig. 1b). These acquisitions identified the regions of the cerebral cortex that responded more robustly to the intention to move the left and right lower limbs. We integrated this information with anatomical constraints, to define the optimal positioning of the two ECoG recording implants that aim to decode left and right lower limb movements. The location of both implants was uploaded onto a neuronavigation system to establish the pre-operative planning of the neurosurgical intervention.

Under general anaesthesia, a bicoronal incision of the scalp was performed to enable two circular-shaped craniotomies over the planned locations of the left and right hemispheres, using a tailor-made circular trephine that matched the diameter of the implants. We then replaced the bone flaps with the two implantable recording devices, before closing the scalp.

The paddle lead was positioned over the dorsal root entry zones of the lumbar spinal cord during the STIMO clinical trial. The optimal position of the lead was identified using a personalized model of the spine elaborated from high-resolution structural imaging[5] (Fig. 1c). The final location was optimized intra-operatively on the basis of electrophysiological recordings[4,5]. The implantable pulse generator, which was connected to the lead, was inserted into an abdominal subcutaneous pocket.

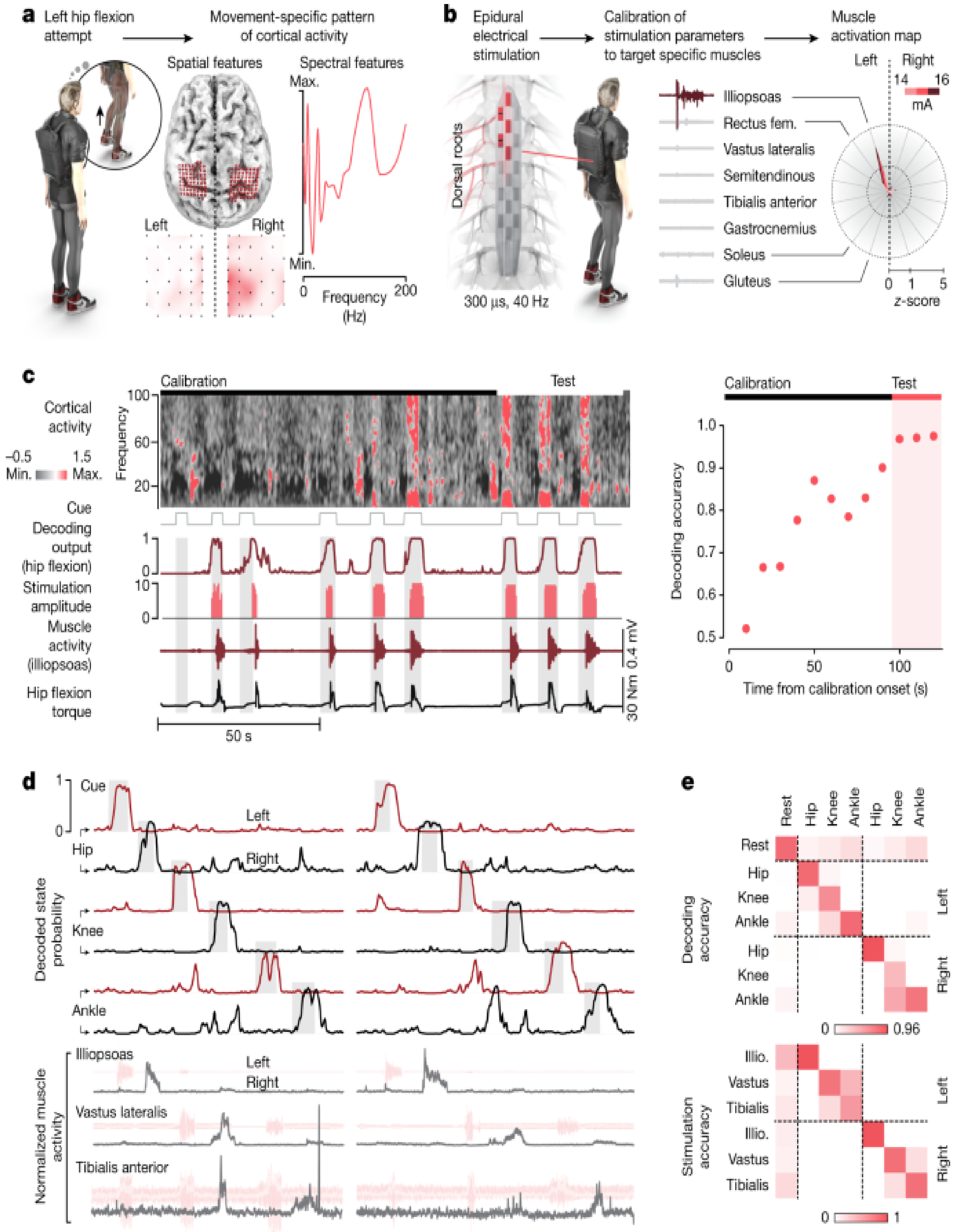
The participant was discharged 24 h after each neurosurgical intervention.

Setup of cortical and spinal implants

The calibration of the BSI required two independent procedures to select the features of ECoG recordings that discriminate the intention to move, and to configure stimulation programs that modulate specific ensembles of lower limb muscles.

The first procedure consisted of extracting the spatial, spectral and temporal features of ECoG signals that were linked to the intention to mobilize each joint of both lower limbs. For this purpose, the participant was asked to attempt hip, knee and ankle movements of the left and right sides in a seated position, during which ECoG signals were recorded concurrently. This mapping enabled the identification of the electrodes, spectral features and temporal windows that captured the larger amount of movement-related information[4,21,22,23,24] (Fig. 2a and Extended Data Fig. 2). The electrodes that measured neural signals correlating with leg movements were located on the most medial aspect of the implant, rostral to the central sulcus, as expected on the basis of pre-operative magnetoencephalographic recordings. The spatial distribution of these electrodes followed a somatotopy that enabled the accurate discrimination of hip, knee and ankle movements (Extended Data Fig. 2c). On the other hand, upper limb-related movements coincided with the modulation of ECoG signals measured through electrodes located on the lateral aspect of the implant (Extended Data Fig. 2f). Movement-related information was contained over the entire range of beta and gamma frequency bands of the ECoG signals (Fig. 2a and Extended Data Fig. 2g). This procedure enabled us to configure the implants with optimal features to enable the participant to operate the BSI (Extended Data Fig. 2f,g).

Fig. 2: Calibration of the BSI.



a, Identification of the spatial and spectral distributions of ECoG feature weights related to attempted left hip flexions. **b**, Calibration of anode/cathode configurations and stimulation parameters (frequency, range

of amplitudes) to elicit left hip flexions, including electromyographic signals from lower limb muscles. The polar plot reports the relative amplitude of muscle responses for the optimal configuration to target left hip flexors over the range of functional stimulation amplitudes (300 µs, 40 Hz, 14–16 mA). **c**, Online calibration of the BSI to enable volitional hip flexion in a seated position. Representative sequence reporting spectrogram, decoding probability and proportional modulation of stimulation amplitudes together with the resulting muscle activity and torque. The plot reports the convergence of the model over time, reaching 97 ± 0.4% after 90 s. **d**, Similar representations after the calibration of the BSI to enable the control over hip, knee and ankle joints of the lower limbs. **e**, Confusion matrices reporting the decoding accuracy for each joint (74 ± 7% s.e.m.) and the accuracy of the stimulation for each targeted muscle group (83 ± 6% s.e.m.).

The second procedure consisted of configuring stimulation programs (Fig. 2b). Epidural electrical stimulation of the spinal cord can modulate specific ensembles of motor pools through the recruitment of the dorsal root entry zones projecting to the spinal cord regions wherein these motor pools reside[9,25]. In turn, optimized configurations of anodes and cathodes can steer electric fields towards specific subsets of dorsal root entry zones to modulate well-defined ensembles of motor neuron pools[5,9,25]. This physiological principle enables the regulation of extension and flexion movements from each joint. We leveraged this principle to configure a library of targeted epidural electrical stimulation programs that mobilized the hip, knee and ankle joints from both sides. Concretely, we configured combinations of anodes and cathodes, stimulation frequencies and amplitudes to steer electrical currents to achieve gradual control over the activity of the targeted muscle groups (Extended Data Fig. 2b–d).

Adaptive online calibration of the BSI

We then leveraged the configurations of cortical and spinal implants to calibrate the BSI on the basis of a recursive, exponentially weighted Aksenova/Markov-switching multilinear algorithm that linked ECoG signals to the control of epidural electrical stimulation parameters (Extended Data Fig. 1).

The algorithm was designed to generate two separate predictions. First, a gating model calculated the probability of the intention to move a specific joint. Second, an independent multilinear model predicted the amplitude and direction of the intended movement. The adaptive properties of the algorithm enabled online, incremental parametrization of the models throughout the period of calibration. A hidden Markov model ensured the stability and robustness of the predictions[26].

We then translated the predictions of the algorithm into an analogue controller that adjusted the amplitude of joint-specific stimulation commands. These updated commands were delivered to the implantable pulse generator every 300 ms.

As early as the first session after neurosurgical intervention, the algorithm calibrated a BSI that enabled the participant to control the relative flexion of the left and right hips of an avatar projected on a screen (Supplementary Video 2). We then integrated the analogue control over the stimulation amplitude to the algorithm. From a lying position, within less than two minutes the participant was able to control the activity of hip muscles to generate a torque with an accuracy of 97% (Fig. 2c).

We then expanded this BSI framework to enable the participant to control the relative amplitude of hip, knee and ankle joints bilaterally along with the resting state, amounting to a total of seven states. Using this proportional BSI combining seven states, the participant achieved gradual control over the movement of each joint bilaterally with an accuracy of 74 ± 7%, whereas the chance level was limited to 14% (Fig. 2d,e). The latency of the decoder was as low as 1.1 s (±0.15 s s.e.m.) for the seven states.

These early sessions validated the procedure for the rapid, robust and accurate calibration of a BSI operating over multiple dimensions.

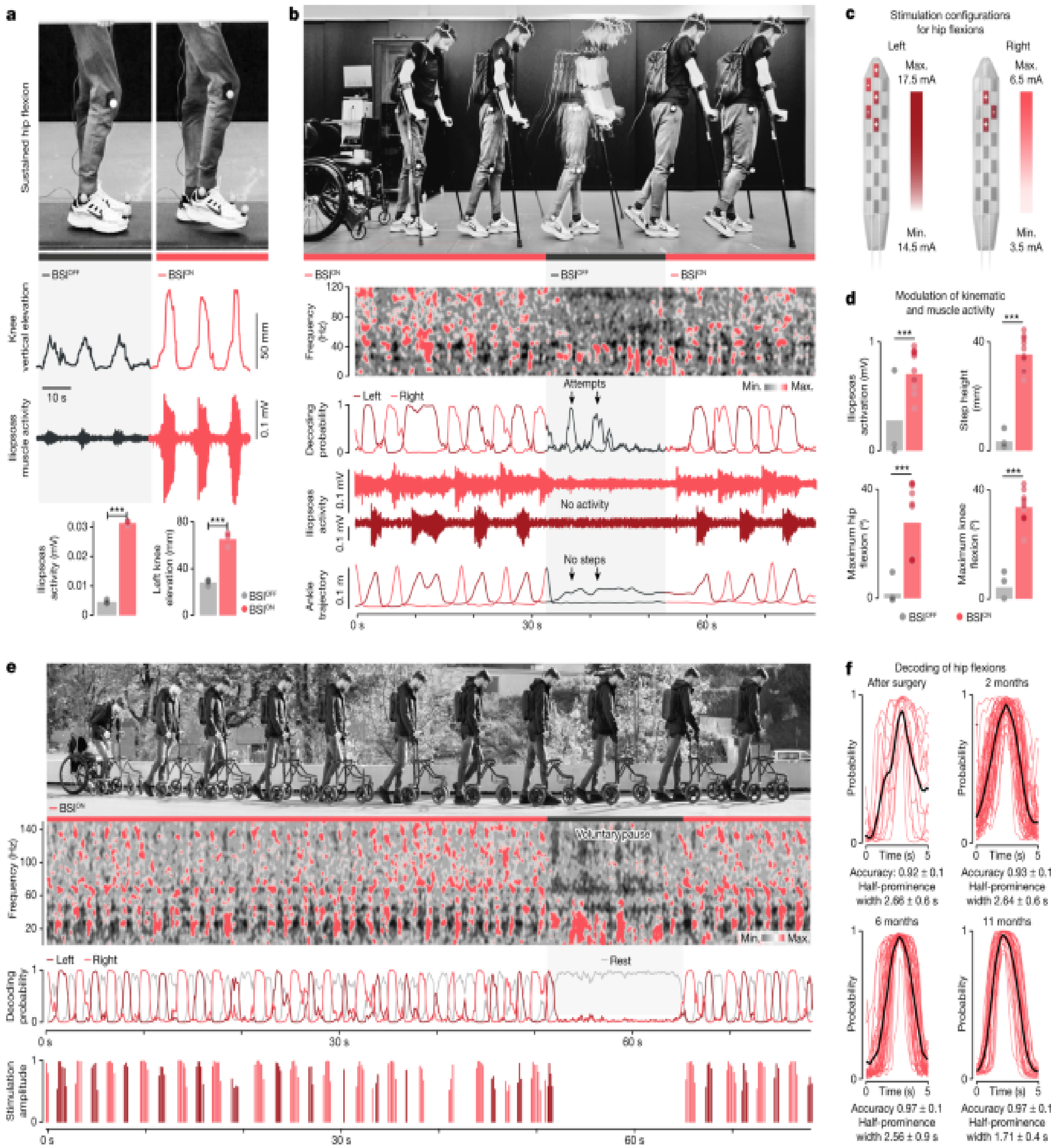
Immediate recovery of natural walking

We next asked whether this procedure supports the calibration of a BSI that restores natural control of walking.

Walking involves well-defined sequences of muscle activation patterns that support weight acceptance, propulsion and swing of the left and right lower limbs. These sequences coincide with the activation of motor pools located within well-segregated regions of the lumbosacral spinal cord[4,27]. Therefore, we selected the stimulation programs within the library that targeted muscles associated with weight acceptance, propulsion and swing functions, and linked these programs to decoding probabilities. We calibrated the BSI to enable the participant to control the relative amplitude of stimulation programs for weight acceptance and swing functions.

We first tested this BSI during voluntary elevations of the foot while standing. After only 5 min of calibration, the BSI supported continuous control over the activity of hip flexor muscles, which enabled the participant to achieve a fivefold increase in muscle activity compared to attempts without the BSI (Fig. 3a).

Fig. 3: The BSI restores natural control of walking.



a, Attempts to perform voluntary hip flexions without and with the BSI, including photographs, vertical elevation of the knee and hip flexor muscle activity. Bar plots report the mean values for these measurements. ($n = 3$ attempts per condition, unpaired one-tailed t-test, $***P < 0.001$.) **b**, Chronophotography during walking with the BSI turned on, off and then on again. Note the two decoded attempts that do not lead to muscle activity nor the execution of steps. **c**, Range of stimulation amplitude during walking. **d**, Bar plots reporting mean values of kinematic and muscle activity parameters

during walking with the BSI turned off and on, ($n = 3$ and 8 attempts for BSI^{OFF} and BSI^{ON}, respectively, unpaired one-tailed *t*-test, $^{***}P < 0.001$, P(iliopsoas activation) $= 3.4 \times 10^{-4}$, P(step height) $= 5.1 \times 10^{-10}$, P(hip angle) $= 2.7 \times 10^{-5}$, P(knee angle) $= 1.6 \times 10^{-9}$). **e**, Chronophotography of standing (voluntary pause) and walking with the BSI outdoors. The spectrogram, probabilities of left and right steps and modulation of stimulation amplitudes illustrate the robustness of the performance and absence of false-positive detections during the voluntary pause. **f**, Plots report the probability of right hip flexions over consecutive steps measured during the first session after the neurosurgical implantation ($n = 13$ steps, accuracy $= 0.92 \pm 0.1$ s.d., $w = 2.66 \pm 0.6$ s s.d.), and at 2 ($n = 46$ steps, accuracy $= 0.93 \pm 0.1$ s.d., $w = 2.64 \pm 0.6$ s s.d.), 6 ($n = 41$ steps, accuracy $=$ 0.97 ± 0.1 s.d., $w = 2.56 \pm 0.9$ s s.d.) and 11 months ($n = 29$ steps, accuracy $=$ 0.97 ± 0.1 s.d., $w = 1.71 \pm 0.4$ s s.d.) after the first activation of the BSI using updated models (Extended Data Fig. 5).

We provided the same configuration to support walking with crutches. The BSI enabled continuous, intuitive and robust control of walking (Fig. 3b). When the BSI was turned off, the participant instantly lost the ability to perform any step, despite detected attempts to walk from the modulation of cortical activity. Walking resumed as soon as the BSI was turned back on. The participant was able to decide whether to initiate stepping, walk continuously, stop or stand quietly without the detection of false-positives that would impair standing performance (Extended Data Fig. 3). Indeed, Berg Balance Scale assessments revealed that the BSI did not impair, and even slightly improved, overall balance abilities (Extended Data Fig. 3c).

The participant reported that the BSI enabled a natural control over his movements during walking (Supplementary Video 2). We aimed to capture this subjective perception with quantified outcomes. For this purpose, we applied a principal component analysis to whole-body kinematics and muscle activity collected during walking on a treadmill with the BSI or with the same stimulation programs controlled in a closed loop on the basis of motion sensors attached to the feet. Compared to stimulation alone, the BSI enabled walking with gait features that were markedly closer to those quantified in healthy individuals (Extended Data Fig. 4a). The BSI ensured a continuous link between the intended movement and the modulation of

stimulation protocols, which translated into the ability to walk overground independently with crutches. When the intended movements were detected from the motion sensors, the participant reported a frequent temporal mismatch between the detections and his intentions, which impaired his ability to walk under these conditions (Extended Data Fig. 4b).

Navigation over complex terrain

We next aimed to investigate whether the BSI could enable an intuitive and natural control over complex activities of daily living that were not possible without the BSI.

When the participant enroled in STIMO, seven years after his accident, he was not able to walk independently. Completion of this clinical trial enabled him to regain basic walking when stimulation was turned on, albeit that this recovery required compensatory strategies to trigger the sequences of stimulation based on heel elevations. He also recovered partial mobility without stimulation. However, he experienced difficulties transitioning from standing to walking and stopping, and could only walk over flat surfaces. Moreover, he was not able to adjust lower limb movements to progress over ramps, overcome obstacles or climb up staircases—as is necessary to support mobility in everyday life.

To demonstrate that the BSI remedied these limitations, we designed a succession of models that emulated the conditions underlying these activities of daily living.

We first asked whether the participant was able to walk on steep terrain requiring adaptive modulation of the amplitude of muscle activity. With the BSI, the participant climbed up and down a steep ramp with ease, performing this task two times faster than without stimulation. The BSI also enabled high step clearance, as necessary to climb over a succession of stairs, negotiate obstacles and traverse changing terrains (Extended Data Fig. 4c,d). All these tasks were performed with the same BSI configuration, which proved highly reliable to support a broad variety of tasks with widely different constraints (Extended Data Fig. 4c,d).

Long-term stability of the BSI

We next sought to assess the stability of the BSI. For this purpose, we quantified the stability of cortical signals and decoders over time, and the need to adjust stimulation programs.

After a transitory one-month period, during which cortical signals exhibited modest changes in the spectral content of the different frequency bands, ECoG signals remained stable over the following months (Extended Data Fig. 5a). The decrease in the spectral power was limited to 0.03 dB per day on average. This stability enabled robust performance. For example, we found that the same decoder enabled the participant to achieve gradual control over six joints despite a two-month interval between both sessions (Extended Data Fig. 6). We leveraged this robustness during neurorehabilitation, as we only recalibrated the BSI when deemed necessary by the participant and/or physiotherapists to promote the best possible functional performance. Despite these recalibrations, the features of the decoders remained remarkably stable over time (Extended Data Fig. 5b). Indeed, signal quality and decoding accuracy during walking has remained globally unchanged over nearly one year of use (Fig. 3f and Extended Data Fig. 5d). Whereas cortical features remained stable over time, we detected a progressive reinforcement of their modulation depth, which revealed gradual improvements in the ability of the participant to modulate his cortical activity when operating the BSI (Extended Data Fig. 5e).

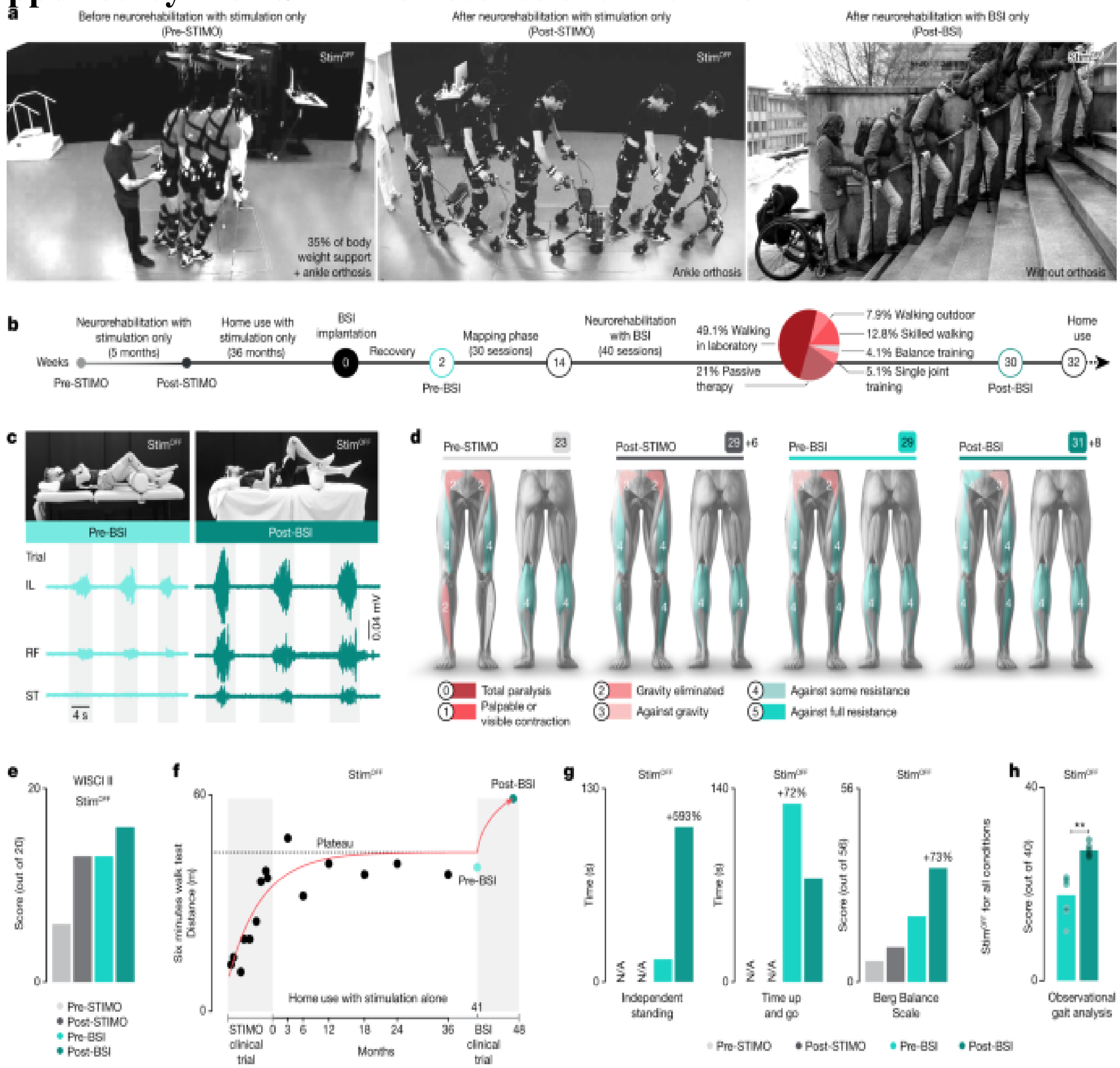
The library of stimulation programs showed the same stability. The optimal range of stimulation amplitudes was dependent on the specific configuration of electrodes and targeted muscles (Extended Data Fig. 5c). However, these ranges of stimulation amplitudes have remained stable over one year of use, and stimulation thresholds did not change over time.

Neurological recovery

The clinical study was designed to investigate whether neurorehabilitation supported by the BSI further improves neurological recovery (Fig. 4a). Before enroling in STIMO-BSI, the participant had completed the clinical trial STIMO, which enabled him to regain volitional control over previously

paralysed muscles and to improve his standing and walking functions. However, after three years of regular training with stimulation only, he had reached a plateau of recovery (Fig. 4d–f).

Fig. 4: Neurological improvements following neurorehabilitation supported by the BSI in the absence of stimulation.



a, Chronophotography illustrating the walking ability of the participant without any stimulation before enroling in the STIMO clinical trial (pre-STIMO), after its completion (post-STIMO) and after completion of the STIMO-BSI clinical trial (post-BSI). **b**, Timeline of the two clinical trials, including a pie chart reporting the time during which the various types of neurorehabilitation exercises were practised, as well as the home use of the

BSI. **c**, Photographs showing the maximal hip flexion and associated flexor muscle activity before and after neurorehabilitation. **d**, Changes in lower limb motor scores over the course of both clinical trials. **e**, Plots reporting improvements in WISCI II scores over the course of both clinical trials. Neurorehabilitation supported by the BSI restored the capacity to walk over 10 m with crutches without any assistance or stimulation. **f**–**h**, Plots reporting quantifications of the 6 min walk test (**f**), weight-bearing capacity, time up and go, Berg Balance Scale (**g**) and observational gait analysis (**h**) (each dot refers to scores from a physiotherapist ($n = 6$, paired one-tailed t-test; $^{**}P = 0.002$). N/A, not available.

The participant completed 40 sessions of neurorehabilitation (Fig. 4b). which involved walking with BSI, single-joint movements with BSI, balance with BSI and standard physiotherapy. Because impairments were more pronounced in hip flexor muscles, we primarily focused the training exercises and BSI configurations on the control of these muscles.

This neurorehabilitation programme mediated pronounced improvement in the volitional control of hip flexor muscles and associated hip flexion movements without stimulation (Fig. 4c). This recovery correlated with gains in sensory (4 points in light touch sensory score) and motor scores (Fig. 4d), and enhanced standing and walking capacities that were captured in an increase in WISCI II scores from 6 before STIMO to 16 after STIMO-BSI (Fig. 4e). Concretely, the participant exhibited improvements in all the conventional clinical assessments, such as the six-minute walk test, weight-bearing capacities, timed up and go, Berg Balance Scale and walking quality assessed using the observational gait analysis scale[28] by physiotherapists blinded to the study (Fig. 4d–h and Extended Data Table 3). These improvements without stimulation translated into a meaningful increase in quality of life, such as walking independently around the house, transiting in and out of a car or drinking a beverage with friends standing at a bar (Supplementary Video 3).

Integration of the BSI in daily life

The BSI enhanced the standing and walking capacities of the participant, which compelled us to develop a BSI framework for independent use at

home.

We designed a system that could be operated by the participant without any assistance. This system includes a walker equipped with an integrated case that embeds all the components of the BSI (Extended Data Fig. 7). A tactile-based interface enables the participant to interact with the tailored software to launch an activity, verify the placement of the headset and adjust the minimum and maximum amplitudes of stimulation programs. The configuration of the hardware and software is completed with minimal user inputs within less than 5 min, after which the participant can leverage the BSI for neurorehabilitation or to support activities of daily living (Supplementary Video 4). The participant used the system regularly over the course of 7 months with stable decoding performance (Extended Data Fig. 7c). This home use translated into a broad increase in the perceived benefits by the participant, as quantified by the Psychosocial Impact of Assistive Devices Scale (PIADS) questionnaire (Extended Data Table 4). Security, skilfulness and the ability to participate were ranked with the maximum possible gains in this questionnaire.

Discussion

We conceived a wireless, digital bridge between the brain and spinal cord that restored natural control over lower limb movements to stand and walk on complex terrains after paralysis due to a spinal cord injury. Moreover, neurorehabilitation mediated neurological improvements that persisted even when the bridge was switched off.

The validation of this digital bridge was restricted to a single individual with severe but partial damage of the spinal cord, and it therefore remains unclear whether the BSI will be applicable to other injury locations and severities. However, several observations suggest that this approach will be applicable to a broad population of individuals with paralysis. First, the physiological principles underlying targeted epidural electrical stimulation of the spinal cord have now been validated in nine out nine treated individuals with incomplete[4] and complete[5] injuries[6]. Second, we developed procedures that supported straightforward, rapid and stable calibration of the link between cortical activity and stimulation programs, enabling the participant to

operate the BSI at home without supervision. Third, a comparable robustness and stability of this computational and technological brain decoding framework has now been observed in two additional individuals with tetraplegia[3,26,29]. Although the previous experience of the participant with the stimulation accelerated the configuration of the BSI, we do not anticipate major impediments to implement a BSI in new individuals. Indeed, we were able to configure stimulation programs that restored stepping within one day in three participants with complete sensorimotor paralysis[5].

The delivery of epidural electrical stimulation over the lumbar spinal cord has enabled many individuals with spinal cord injury to regain adaptive control over the activity of otherwise paralysed muscles. This recovery has been documented in various independent studies, including in participants with complete sensorimotor paralysis[5,6,30,31,32,33,34]. These observations indicate that anatomically intact, yet functionally silent pathways from the brain can modulate the impact of epidural electrical stimulation on the activity of the spinal cord below the injury. However, these studies also acknowledge a series of limitations. First, stimulation parameters must be fine-tuned on the basis of the targeted muscle or desired motor function. Second, the onset of the stimulation must be precisely synchronized with the motor intention. Third, finely graded control over the activity of muscles requires modulating the amplitude of the stimulation. The BSI remedies these three limitations. In this scenario, the residual and prosthetic pathways converge on the same neurons below the injury, enabling graded and sustained control over the activity of muscles. This cooperation probably plays an essential role in the reorganization of neuronal pathways that mediate neurological recovery in response to neurorehabilitation with the BSI. Comparable observations have been reported when electroencephalographic signals were coupled to an exoskeleton or functional electrical stimulation of muscles during gait rehabilitation in people with spinal cord injury[22,24,35]. However, the poor quality of electroencephalographic signals in mobile conditions, combined with the impracticality of this technological framework, are an impediment to the clinical implementation of these non-invasive strategies.

Neurorehabilitation supported by the digital bridge mediated additional neurological improvements after three years of stable performance, despite continued use of epidural electrical stimulation at home. These improvements primarily took place in the control of hip muscles, which was the main target of brain-controlled stimulation programs during neurorehabilitation. Although focused on one muscle group, this neurological recovery translated into the ability to lift the leg against gravity without stimulation. This recovery supported independent walking with crutches.

In preclinical models, neurorehabilitation supported by a digital bridge triggered superior recovery compared to epidural electrical stimulation alone[15]. Brain-controlled neuromuscular stimulation also mediated durable functional improvements of the engaged muscles after stroke[36] and spinal cord injury[22,24,37]. As the participant had previously reached a plateau of recovery after intensive rehabilitation using spinal cord stimulation alone, it is reasonable to assume that the BSI triggered a reorganization of neuronal pathways that was responsible for the additional neurological recovery. These results suggest that establishing a continuous link between the brain and spinal cord promotes the reorganization of residual neuronal pathways that link these two regions under normal physiological conditions[38,39,40,41]. Expanding the concept of a digital bridge to the cervical spinal cord may also restore arm and hand movements after spinal cord injury[42] and stroke[43]. However, it is important to appreciate that the relative amount of neurological recovery will necessarily correlate with the severity of the lesion.

Scaling up this digital bridge will require several developments. First, the practical utilization of the cortical implant will necessitate miniaturization of the base station, computing unit and unnoticeable antennas. Compressive sensing and dynamic adjustment of sampled electrodes and features could further reduce cortical device footprint. Second, the spinal implant must be endowed with ultrafast communication capabilities, versatile stimulation parameters and direct wireless control from the wearable computing unit. Finally, the cortical and spinal implants could be controlled by a single low-power integrated circuit embedding a neuromorphic processor with self-calibration capability that autonomously translates cortical activity into

updates of stimulation programs. Although these developments require time and resources, we do not anticipate technical hurdles to realize this transition.

The concept of a digital bridge between the brain and spinal cord augurs a new era in the treatment of motor deficits due to neurological disorders.

Methods

Study design and participant

All experiments were carried out as part of the ongoing clinical feasibility study STIMO-BSI ('Brain-controlled Spinal Cord Stimulation in Patients With Spinal Cord Injury'), which investigates the safety and preliminary efficacy of brain-controlled spinal cord stimulation after spinal cord injury (clinicaltrials.gov, NCT04632290). All surgical and experimental procedures were performed at the Lausanne University Hospital (CHUV) except for the magnetoencephalography experiments, which were performed at the CEA Clinatec facilities (Grenoble). The study involved functional assessments before implantation of the cortical devices, the neurosurgical procedure, a 6 week period during which various decoders were calibrated and spinal cord stimulation libraries were established, and a 15 week period of neurorehabilitation with physiotherapists. which amounted to a total of 40 sessions lasting one to three hours. The neurorehabilitation programme was personalized on the basis of the participant's improvements. At the end of the neurorehabilitation period, the participant exited the active participation phase of the clinical trial, and was offered the opportunity to continue using the BSI at home. The participant is currently being followed up on a regular basis by the study team until the end of the three-year study extension phase of home use of the system.

Before enrolment in the STIMO-BSI study, the participant had completed the clinical protocol STIMO ('STIMO: Epidural Electrical Simulation (EES) With Robot-assisted Rehabilitation in Patients With Spinal Cord Injury', NCT02936453) during which a spinal cord stimulation system had been implanted and he had completed a five-month intensive neurorehabilitation

programme supported by EES, followed by a two-year period of independent use at home.

Additionally, one year before joining the STIMO-BSI trial, the participant underwent a surgical procedure with: (1) talonavicular arthrodesis, transfer of the toe extensors to the peroneus tertius; and transfer of the tibialis posterior to the tibialis anterior and extensor digitorum longus; and (2) tenotomy of all long toe flexors and interphalangeal arthrodesis of the hallux. Both were performed bilaterally and could have impacted the reliability of the long toe extensor motor scores before and during the study due to change in spasticity and mechanical properties of the joint. Therefore, we decided not to report the long toe extensor motor score in our analysis.

Pre-operative magnetoencephalography

Before entering the STIMO-BSI clinical trial, the participant was already implanted with a spinal cord system that was not MRI compatible. Therefore, we were not able to perform anatomical or functional MRI of the brain. Magnetoencephalography (MEG) is less sensitive to anatomical imprecisions for source reconstruction compared to electroencephalography (EEG)[44]. Therefore, we decided to use MEG to map the activity correlated to the limb motor intentions.

Before the neurosurgical procedure to place the cortical implants, MEG activity was measured in a magnetically shielded room using a 306-channel whole-scalp array (204 planar gradiometers and 102 magnetometers) from the Elekta Neuromag system (Elekta Neuromag). Electrocardiogram and electro-oculogram were recorded simultaneously. The recording sampling rate was 1,000 Hz. Continuous head position indicator signals were recorded during the experiments to track the head movements of the subject. Before experimentation, a three-dimensional (3D) digitization system (Isotrak II, Polhemus) was used to localize anatomical fiducial points for later coregistration with head computerized tomography (CT). Temporal signal space separation (tSSS) was applied to reduce the noise in the MEG data using MaxFilter v.3.0 software (Elekta). First, manual review of raw data enabled the marking of bad channels. Second, the tSSS filter was applied using head movement compensation and automatic bad channel correction.

The main parameters were kept to default (tSSS correlation threshold of 0.98, orders of expansion for ‘in’ and ‘out’ components of signal set to 8 and 3, respectively, and a 10 s time buffer). Notch filtering at 50 Hz and harmonics (100 Hz, 150 Hz, 200 Hz and 250 Hz) were also applied to remove power line contamination. Stereotypical artefacts (cardiac, ocular) were identified by independent components analysis using MNE-Python software[45] and rejected on visual screening (Infomax method, calculated separately for magnetometers and gradiometers using 64 components). The head, skull and cortex geometries were calculated from CT scan using the MRI segmentation routine included in the Brainstorm software[46], followed by calculation of the head model using the overlapping spheres method. A 3D inversion kernel was calculated using Brainstorm implementation of the Minimum Norm Imaging method with default parameters. It enabled the reconstruction of cleaned raw data at brain source level for the subsequent calculations. Finally, the MSA method[47] was used to reconstruct the brain activity related to wrist, hip and ankle motor attempts. To estimate task-specific brain activations, MSA uses cross-validated, shifted, multiple Pearson correlation, calculated from the time–frequency transformed brain signal and the binary signal of stimuli. 3D snapshots of these activations at their maximum were exported as DICOM in the original CT scan frame of reference for use in neuronavigation tools. For rendering, we manually segmented the brain from the patient pre-operative CT scan using Slicer (slicer.org) and used Blender for rendering. We recoloured the MEG signal with a red colour ramp then superposed it with the 3D render of the brain.

Electrocorticography WIMAGINE devices

The WIMAGINE implantable recording system was designed for bilateral epidural implantation over the sensorimotor cortex[20]. Electronic components were housed in a titanium case (50 mm diameter, 7–12 mm thick and a convex external face). An array of 64 platinum–iridium (90:10) recording electrodes for epidural ECoG (2 mm in diameter, 4–4.5 mm pitch) and five reference electrodes were located on the flat inner face of the device. The ECoG data were recorded thanks to an application-specific integrated circuit[48] that enabled multi-channel amplification and digitalization with an input referred noise of less than 0.7 μV root mean square in the 0.5 Hz–300 Hz range. Data were radio-emitted through an ultrahigh frequency antenna

(402–405 MHz). Power was supplied remotely through a 13.56 MHz inductive high-frequency antenna. Both antennas were embedded in a silicone flap extending on the subcutaneous space. To ensure signal stability at high frequency (586 Hz), with regard to the limited bandwidth, 32 contacts out of the 64 were used for each implant. The wireless connection used two external antennas held in front of the recorders by a custom-designed headset. The technical specifications of the device are reported in Extended Data Table 1.

Neurosurgical procedure

The surgery was performed under general anaesthesia. A neuronavigation station (StealthStation, Medtronic) was used to locate the centre of the craniotomies on each hemisphere. The anatomical and functional information obtained from MEG and CT scan imaging enabled the selection of the entry points to maximize the coverage of the leg region of the sensorimotor cortex, while ensuring a safe margin from the sagittal sinus area. Following a coronal incision, two circular craniotomies of 5 cm in diameter were performed using a custom-made trephine. The bone flaps were removed to expose the dura mater. The two WIMAGINE implants were placed over the dura, and then carefully suspended and secured with non-resorbable sutures. The skin was then sutured over the implants. The participant was discharged the next day. The calibration phase was initiated after a rest period of two weeks.

Decoder architecture

ECoG data were collected from 32 channels per implant at an acquisition frequency of 586 Hz. The signals were band-pass filtered between 1 Hz and 300 Hz. The data were streamed through the fieldtrip toolbox to a custom-made decoding software running in the Matlab Runtime Environment (Mathworks).

To decode the intention to perform lower limb movements, we implemented a variant of the recursive exponentially weighted Markov-switching multilinear model (REW-MSLM) algorithm that we had previously developed to decode upper limb movements[26]. REW-MSLM is a mixture of

multilinear algorithms form experts. It consists of a hidden Markov model (HMM) classifier, called ‘gate’, and a set of independent regression models, called ‘experts’. Each expert is generally dedicated to the control of a group of degrees of freedom, a specific limb or movement (for example, joint movement). The HMM-based classifier predicts the probability of such specific limb or movement activation (states) associated with a particular expert. The resulting decoder output results from soft mixing of expert predictions according to estimated probabilities.

The gate HMM-based classifier predicts the state and assumes that the state sequence $Z(t)$ follows a first-order Markov chain hypothesis. Consequently, the probability of a state at each time step depends on the combination of the previous state and the newly acquired ECoG data. The HMM-based classifier is composed of an emission and transition probability model. At each time step, the emission probability is estimated from the observations of ECoG signals independently from the sequence of the state. In the current study, a linear discriminative classifier was used for the emission probability model. For K states/classes (in our case, $K = 7$ states, rest, hip, knee, ankle, bilaterally), the classifier output was computed as follows:

$${{\bf{d}}}^{{\rm{g}}{\rm{a}}{\rm{t}}{\rm{e}}}(t)={\beta }^{{\rm{g}}{\rm{a}}{\rm{t}}{\rm{e}}}X(t)+{{\bf{b}}}^{{\rm{g}}{\rm{a}}{\rm{t}}{\rm{e}}},$$

$${{\bf{d}}}^{{\rm{g}}{\rm{a}}{\rm{t}}{\rm{e}}}(t)=({d}_{1}^{{\rm{e}}{\rm{m}}{\rm{i}}{\rm{s}}{\rm{s}}{\rm{i}}{\rm{o}}{\rm{n}}}\,(t),\ldots ,{d}_{K}^{{\rm{e}}{\rm{m}}{\rm{i}}{\rm{s}}{\rm{s}}{\rm{i}}{\rm{o}}{\rm{n}}}\,(t)).$$

Here $\beta^{\rm gate}$ and $\mathrm{b}^{\rm gate}$ are matrices of coefficients and bias of linear discriminative classifiers. Then, the emission probability vector ${\alpha }^{{\rm{e}}{\rm{m}}{\rm{i}}{\rm{s}}{\rm{s}}{\rm{i}}{\rm{o}}{\rm{n}}}\,(t)\,=$ $\left({\alpha }_{1}^{{\rm{emission}}}\,(t),\ldots ,{\alpha }_{K}^{{\rm{emission}}}\,(t)\right)$ is obtained from the classifier output $\mathrm{d}^{\rm gate}(t)$ through the softmax normalization:

$${\alpha }_{k}^{{\rm{emission}}}\,(t)=\frac{{{\rm{e}}}^{{d}_{k}^{{\rm{gate}}}(t)}}{\mathop{\sum }\limits_{i=1}^{K}{{\rm{e}}}^{{d}_{i}^{{\rm{gate}}}(t)}}.$$

Finally, the emission probabilities are weighted by the HMM state transition probabilities matrix T, where T is a K-by-K matrix with coefficients defined by the cumulative number of transitions between cued states.

$$\hat{\alpha }(t)=\frac{{\alpha }^{{\rm{e}}{\rm{m}}{\rm{i}}{\rm{s}}{\rm{s}}{\rm{i}}{\rm{o}}{\rm{n}}}\,(t)T\hat{\alpha }(t-1)}{{\alpha }^{{\rm{e}}{\rm{m}}{\rm{i}}{\rm{s}}{\rm{s}}{\rm{i}}{\rm{o}}{\rm{n}}}\,(t)T\hat{\alpha }(t-1)}.$$

The most probable state $\hat{Z}\left(t\right)$ sequence may be issued by maximizing the state probability $\hat{\alpha }(t)$ at step t. The state probability vector may be used to mix the decoder experts, or may be considered as one of the decoder outputs.

For experts, a multilinear regression model was used:

$${\phi }_{k}(t)={\beta }_{k}^{{\rm{e}}{\rm{x}}{\rm{p}}{\rm{e}}{\rm{r}}{\rm{t}}}X(t)+{{\bf{b}}}_{k}^{{\rm{e}}{\rm{x}}{\rm{p}}{\rm{e}}{\rm{r}}{\rm{t}}},$$

where ${\beta }_{k}^{{\rm{expert}}}$ and ${b}_{k}^{{\rm{expert}}}$ summarize coefficients of kth expert, $k \in [1, K]$. Finally, the mixture of expert output $U(t)$ is computed from expert predictions $\varphi_k(t)$ and estimated probabilities ${\hat{\alpha }}_{k}\left(t\right)$ from the following equation: ${U}_{k}\left(t\right)={\phi }_{k}\left(t\right)\times {\hat{\alpha }}_{k}\left(t\right)\times \prod _{i\ne k}(1-{\hat{\alpha }}_{i}\left(t\right))$.

From this decoder architecture, we implemented two different control models to drive epidural spinal cord stimulation.

1. (1)

   For six-joint control in static conditions, we implemented a mixture of expert predictions to enable the participant to achieve proportional control over the amplitude of stimulation. $U(t)$ contains the analogue prediction of the relative desired amplitude of joint movement at any given time. The movement of each joint is linked to a specific

stimulation protocol (electrode configuration, frequency and pulse width) defined in the library of stimulation programs (Extended Data Fig. 2b–d), whereas the predictions constituting $U(t)$ are linearly rescaled into amplitude of stimulation (in mA) within a range of predefined values by the experimenter.

2. (2)

For the control of standing and walking in dynamic conditions, we took into consideration that these activities do not require simultaneous control over the amplitude of left and right hip flexions, because left and right steps must not occur at the same time. Therefore, we removed the layer of mixture of experts, and instead implemented a control model that drives stimulation amplitudes from the gate model output. This control model avoids the simultaneous delivery of stimulation over all the joints. Consequently, the amplitude is only modified for one joint at a time. In turn, we used the maximum estimated state probability $\max (\hat{\alpha }\left(t\right))$ to enable the participant to achieve proportional control over the amplitude of stimulation.

Iterative online decoder calibration

REW-MSLM is a closed-loop adaptive decoder. In parallel to the current model use for predictions, the REW-MSLM decoders update their parameters on the basis of new incoming data, which enables the optimization of the parameters of the model in real time throughout the calibration session[26]. The linear emission probability model and expert models were identified using a recursive exponentially weighted N-way partial least square (REW-NPLS) algorithm. This algorithm was specifically designed for incremental and adaptive real-time multilinear decoder learning[49]. The transition matrix was identified by direct state transition counting during the calibration session. The resulting decoder was able to predict mental states as well as continuous movements.

Input features $X(t)$ were computed from the ECoG signals and then fed to the decoder. Epochs ranging from 200 to 500 ms of ECoG signals from the 64 electrodes were created to generate a 100 ms sliding window. The epochs were mapped to the temporal frequency space with complex continuous

wavelet transform (CCWT). The wavelets declined from the Morlet mother wavelet are centred around specific frequencies (2, 5:5:100, 125, 150, 200 Hz). The absolute value of the output of the CCWT was then decimated to obtain 2–5 points along the temporal modality, which defined the epoch. The prediction was computed every 100 ms. During the experiments, the REW-MSLM algorithm recursively updated the experts and gate coefficients every 15 s. The training data consist of 15 s batches of input ECoG features associated to output movement features. The output features are generated according to specific tasks given to the participant to perform motor imagery, including desired mental state for the gate update and a desired continuous movement (if any) for the expert in charge update.

When creating a model from scratch, assistance from the system can be added to the decoder output. This enables the participant to have movements already performed, even though the decoder does not predict correctly. Assistance decreases progressively as the model is calibrated, to eventually leave the participant in full control.

Model update:

Step 1: accumulate raw data and labels d^{gate} and φ over 15 s.

Step 2: compute the corresponding feature vector $X(t)$.

Step 3: perform the partial least square regressions for the gating d^{gate} and experts φ to update the coefficients (β^{gate}, b^{gate}, β^{expert}, b^{expert}).

Step 4: update the transition matrix by adding the number of transitions. $T(i,j) \leq T(i,j) + \text{sum}((Z(t+1),Z(t)) = (i,j))$.

Prediction computation:

Step 1: compute the linear predictions of gate and experts from the current coefficients.

Step 2: apply the exponential normalization and the HMM transition step.

Step 3: mix the prediction from the gating and expert model.

ECoG mapping procedure

To assess the spatial and spectral information in the ECoG signals to discriminate a specific task, we computed the linear regression weights associated with cued motor attempts. To map the features related to lower limb movements (hip, knee, ankle bilaterally) we recorded cortical signals during 57 (±6 s.e.m.) repetitions of each movement attempt cumulating 226 s (±25 s s.e.m.) during each state. The weights generated from this dataset were projected onto the spatial dimension for different frequency bands (0.5–10 Hz, 10–40 Hz, 40–100 Hz and 100–200 Hz) or on the spectral dimension.

Epidural electrical stimulation

The implant to deliver epidural electrical stimulation (Extended Data Table 2) was composed of an ACTIVA RC implantable pulse generator (model 37612, Medtronic) that was interfaced with the Specify Surescan 5-6-5 paddle lead (model 977C190, Medtronic). A dedicated firmware enabled real-time uploads of stimulation tables to control electrical stimulation waveforms[5]. The patient programmer (SPTM, model 09103) was carried within a belt to align its position with the implantable pulse generator. We developed a custom-built stimulation program[5] that sent commands to the patient programmer through a Bluetooth/infrared wireless bridge. The stimulation program enabled the definition of stimulation configurations (cathodes and anodes) and parameters (pulse width, frequencies and amplitude ranges) by an expert user[5]. This chain of software and hardware enabled real-time control of stimulation protocols with a latency inferior to 150 ms (ref. [4]).

Calibration of the library of stimulation programs

Electromyographic (EMG) activity was recorded bilaterally from the iliopsoas, rectus femoris, vastus lateralis, semitendinosus, tibialis anterior, medial gastrocnemius and soleus muscles with wireless bipolar surface electrodes (Delsys Trigno). Each pair of electrodes was placed over the belly of the targeted muscle, aligned longitudinally to muscle fibres. Abrasive paste (Nuprep, 4Weaver) was used for skin preparation to reduce electrode–

skin resistance and improve EMG signal quality. An additional pair of surface-EMG electrodes was placed over the spine, at the thoracolumbar junction, to detect stimulation artefacts, and thus align muscle responses to the onset of stimulation. Continuous EMG signals were sampled at 2 kHz and saved to a desktop computer. EMG signals were band-pass filtered between 20 and 450 Hz. Recruitment curves were performed with pulses of increasing stimulation amplitudes, delivered every second. We implemented a grid search model to explore the different electrode configurations and frequencies to select the configurations of cathodes and anodes to achieve maximum selectivity in the recruitment of the targeted muscle groups[4]. The amplitude of muscle responses was normalized by *z*-scoring over all the configurations. For each period of stimulation, the average absolute value of *z*-score was computed. The *z*-score was then represented in a polar plot.

Decoder calibration procedure

During the calibration of the decoders, the participant received visual cues through a custom-made interface displaying the targeted state with or without the direction of movement. The cues were generated as a pseudo-random sequence with programmable duration (2/4 s per state) or manually by the experimenter. The decoding environment enabled visualization of the duration spent in each state as well as the number of transitions between states. Once the performance of the decoder was judged sufficient by the participant and experimenter, cueing was discontinued and the participant could use the model without further calibration. From day to day, models were updated when deemed necessary. The iterative nature of the implementation facilitated these updates. Typically, the model supporting control over left hip and right hip flexions during walking was trained on the basis of 30 repetitions of each active state, whereas the resting state was predicted from 3 min of ECoG data that were acquired while the participant was performing unspecific hand and trunk movements, as well as talking to ensure robustness of the predictions.

Decoding accuracy quantification

The accuracy of decoding predictions was quantified by computing the normalized cross-correlation between the decoded state $\hat{Z}$ and the

cued state Z after delay compensation:

$${\rm{Decoding}}\;{\rm{accuracy}}=\frac{1}{\sum _{t}Z\left(t\right)}\sum _{t}Z\left(t-\tau \right)\hat{Z}\left(t\right),$$

where τ corresponds to the time at which the maximum of the cross-correlation between the cued state and the decoded state probability is reached.

Muscle response accuracy quantification

Accuracy of muscle responses was obtained by computing the normalized cross-correlation between the decoded state $\hat{Z}$ and the thresholded EMG envelope, which was obtained with $T = 200$ ms sliding window:

$${\rm{EMG}}({\rm{t}})=\left[\frac{1}{{\rm{T}}},\underset{{\rm{t}}}{\overset{{\rm{t}}+{\rm{T}}}{\int }},|{\rm{zscore}}({\rm{emg}}(\tau))|{\rm{d}}\tau \right] > 1$$

$${\rm{EMG}}\,{\rm{accuracy}}=1\,/\,\sum _{t}{\rm{EMG}}(t)\sum _{t}{\rm{EMG}}(t)\hat{Z}(t).$$

Stepping accuracy quantification

When walking freely, there was no cue to quantify the decoding accuracy. To provide a quantification, we profiled the probability curves that were decoded during walking, and analysed peak value and width of the probability curves associated with left and right hip flexions. We conducted this analysis during walking at different time points, from the first session to sessions that occurred nearly one year after the neurosurgical procedure to place the cortical implant. The mean peak of probabilities, as well as the mean half-width (±s.d.) were calculated for each time point.

ECoG spectrograms

To generate spectrograms of ECoG signals, we applied a continuous wavelet transform with a window of 500 ms and a step size of 100 ms. The difference between the averaged spectrograms from both implants was

computed and normalized by applying *z*-scores over each frequency band. The colour maps of the normalized average spectrograms were scaled between −0.5 and 0.5 or between −0.5 and 1 for visualization.

ECoG signal stability

Signal stability was assessed by quantification of the signal power during the resting state in the different frequency bands[50]. The participant was sitting in his wheelchair with eyes closed while ECoG signals were acquired over 2 min. For each session, a window of 90 s starting 20 s after the onset of recording was selected for analysis. The power spectrum density was estimated using Welch's method. The root mean square was computed over the entire frequency band (0.5–292 Hz). The band power was measured for the following four frequency bands: 0.5–10 Hz, 10–40 Hz, 40–100 Hz and 100–200 Hz. To compensate for the different frequency bandwidths, the obtained band powers were normalized before being converted into dB. The signal-to-noise ratio was calculated for each band as the ratio of the band power versus the noise band power, which was estimated between 250–260 Hz due to the numerical filter. Root mean square, band power and signal-to-noise ratio were finally averaged across all the electrodes.

Feature reinforcement with time

We analysed the reinforcement of cortical features linked to hip flexion attempts by computing the median spectrograms around the flexion cues in four different time periods to gather 100 events per period (−2 s to +2 s around the event). For each electrode, we computed the standard deviation of the spectrograms over all frequencies during the 4 s surrounding the events. We performed a linear fit over the 64 electrodes and four time periods.

Walking model stability assessment

To analyse the stability of the walking models, we applied a principal component analysis (PCA) over the coefficients of each gate (idle, left hip flexion, right hip flexion) for each model. The gate vectors were composed of (64 channels × 24 frequencies) coefficients by averaging the temporal

dimension. The PCA was performed over (3 gate × 44 models) samples spanning 4 months of use. Data were represented in the first three components of the PCA. We constructed an ellipsoid of 1,600 data points representing the contour curve that corresponded to a standard deviation of 1.4 for a 3D Gaussian distribution with the covariance and the mean value of each state.

Quantitative gait analysis

EMG activity during walking was acquired bilaterally at 1,259 Hz using 16-channel wireless sensors (Delsys Trigno) placed over the iliopsoas, rectus femoris, vastus lateralis), semitendinosus, tibialis anterior and medial gastrocnemius. Kinematic recordings were acquired using a 3D motion capture system (Vicon Motion Systems). A network of 14 infrared cameras, which covered a $12 \times 4 \times 2.5\ m^3$ workspace, was used to record the motion of markers attached to body landmarks. Data were acquired at a 100 Hz sampling rate using. A PCA was applied over a total of 26 kinematic and EMG parameters that were calculated for each gait cycle, as described previously[4]. The following parameters were included: step length, step height, knee height, knee angle and knee maximum angle, hip angle and hip maximum angle, limb angle, vastus lateralis activation, vastus lateralis stance activation, vastus lateralis swing activation, tibialis anterior activation, tibialis anterior stance activation, tibialis anterior swing activation, rectus femoris activation, rectus femoris stance activation, rectus femoris swing activation, iliopsoas activation, iliopsoas stance activation, iliopsoas swing activation, semitendinosus activation, semitendinosus stance activation, semitendinosus swing activation, medial gastrocnemius activation, medial gastrocnemius stance activation, medial gastrocnemius swing activation. Data were quantified during walking with the BSI and with closed control of stimulation based on motion sensors attached to the feet[5]. These data were compared to identical recordings obtained in healthy individuals. During overground walking with crutches, stepping attempts were detected when the knee angle dropped below 135 degrees with at least 2 s between steps. Steps were considered as failed when the step length was lower than 10 cm.

Observational gait analysis

To analyse gait quality from video recordings, a panel of physiotherapists (*n* = 6), who were blind to experimental conditions and were not involved in STIMO or STIMO-BSI clinical trials, were asked to score different walking trials using items in a validated scoring sheet, which is described in Extended Data Table 3. This scoring sheet pooled items from the validated questionnaires G.A.I.T.[28], SCI-FAI[51], Tinetti Test[52] and ref. [53].

International Standards for Neurological Classification of Spinal Cord Injury

Neurological status was assessed by an experienced neurologist on the basis of the International Standards for Neurological Classification of Spinal Cord Injury (ISNCSCI), a comprehensive clinician-administered neurological examination of residual sensory and motor function quantifying spinal cord injury severity.

Six-minute walk test

Endurance was assessed by the distance covered overground within six minutes with a standard four-wheel walker, but without any external assistance. This test was performed before and at the end of each period of neurorehabilitation of the STIMO and STIMO-BSI. Data were fitted with an exponential curve.

Ten-metre walk test

Walking speed was assessed by a timed ten-metre walk test without any external assistance. The participant was instructed to walk with the preferred assistive device as fast as he could.

Statistical analysis

Individual data points are represented on each figure. Measurements were taken from distinct samples, except for the observational gait analysis for which the expert physiotherapists were independently ranking the same videos. We used paired (when applicable) or unpaired one-tailed *t*-test,

except as otherwise specified, with $\alpha = 0.05$. P values are reported with $***P < 0.001$, $**P < 0.01$ and $*P < 0.05$.

Device explantation

Due to a subcutaneous infection to *Staphylococcus aureus* at the location of the cortical implant located on the right side, the principal investigator decided to explant the device 167 days after implantation. The second implant presented no sign of infection and remained in place, and fully functional. After recovery from the surgery and antibiotics treatment per os, neurorehabilitation and daily use could continue as planned by the protocol. Implantation of a new cortical implant was performed on 9 March 2023.

Ethics statement

The STIMO-BSI study was approved by the Swiss authorities (Swissethics protocol number CER-VD2020-01814, Swissmedic 10000766, EUDAMED CIV-20-07-034126) and was conducted in accordance with the Declaration of Helsinki. The STIMO-BSI study is registered at ClinicalTrials.gov (NCT04632290). The STIMO study was approved by the Swiss authorities (Swissethics protocol number CER-VD PB_2016-00886, Swissmedic 10000234, EUDAMED CIV-16-02-014664), registered at ClinicalTrials.gov (NCT02936453), and was conducted in accordance with the Declaration of Helsinki. The participant signed a written informed consent before to participation. Moreover, the participant gave his consent for the material depicting himself to appear in the contribution and to be published in the journal and associated works without limit on the duration of publication, in any form or medium.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Data presented in this manuscript are available here: https://doi.org/10.5281/zenodo.7680471.

Code availability

Supporting MATLAB scripts are available here: https://doi.org/10.5281/zenodo.7680471.

References

1. Bickenbach, J. et al. *International Perspectives on Spinal Cord Injury: Summary* (WHO, 2013); https://apps.who.int/iris/handle/10665/94192.

2. Ahuja, C. S. et al. Traumatic spinal cord injury. *Nat. Rev. Dis. Primer* **3**, 17018 (2017).

3. Benabid, A. L. et al. An exoskeleton controlled by an epidural wireless brain–machine interface in a tetraplegic patient: a proof-of-concept demonstration. *Lancet Neurol*. https://doi.org/10.1016/S1474-4422(19)30321-7 (2019).

4. Wagner, F. B. et al. Targeted neurotechnology restores walking in humans with spinal cord injury. *Nature* **563**, 65–71 (2018).

5. Rowald, A. et al. Activity-dependent spinal cord neuromodulation rapidly restores trunk and leg motor functions after complete paralysis. *Nat. Med.* **28**, 260–271 (2022).

6. Kathe, C. et al. The neurons that restore walking after paralysis. *Nature* **611**, 540–547 (2022).

7. Arber, S. & Costa, R. M. Connecting neuronal circuits for movement. *Science* **360**, 1403–1404 (2018).

8. Courtine, G. & Sofroniew, M. V. Spinal cord repair: advances in biology and technology. *Nat. Med.* **25**, 898–908 (2019).

9. Capogrosso, M. et al. A computational model for epidural electrical stimulation of spinal sensorimotor circuits. *J. Neurosci.* **33**, 19326–19340 (2013).

10. Wenger, N. et al. Closed-loop neuromodulation of spinal sensorimotor circuits controls refined locomotion after complete spinal cord injury. *Sci. Transl. Med.* **6**, 255ra133 (2014).

11. Wenger, N. et al. Spatiotemporal neuromodulation therapies engaging muscle synergies improve motor control after spinal cord injury. *Nat. Med.* **22**, 138–145 (2016).

12. Moraud, E. M. et al. Mechanisms underlying the neuromodulation of spinal circuits for correcting gait and balance deficits after spinal cord injury. *Neuron* **89**, 814–828 (2016).

13. Capogrosso, M. et al. A brain–spine interface alleviating gait deficits after spinal cord injury in primates. *Nature* **539**, 284–288 (2016).

14. Minev, I. R. et al. Biomaterials. Electronic dura mater for long-term multimodal neural interfaces. *Science* **347**, 159–163 (2015).

15. Bonizzato, M. et al. Brain-controlled modulation of spinal circuits improves recovery from spinal cord injury. *Nat. Commun.* **9**, 3015 (2018).

16. Lorach, H., Charvet, G., Bloch, J. & Courtine, G. Brain–spine interfaces to reverse paralysis. *Natl Sci. Rev.* **9**, nwac009 (2022).

17. Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M. & Shenoy, K. V. High-performance brain-to-text communication via handwriting. *Nature* **593**, 249–254 (2021).

18. Vansteensel, M. J. et al. Fully implanted brain–computer interface in a locked-in patient with ALS. *N. Engl. J. Med.* **375**, 2060–2066 (2016).

19. Moses, D. A. et al. Neuroprosthesis for decoding speech in a paralyzed person with anarthria. *N. Engl. J. Med.* **385**, 217–227 (2021).

20. Mestais, C. S. et al. WIMAGINE: wireless 64-channel ECoG recording implant for long term clinical applications. *IEEE Trans. Neural Syst. Rehabil. Eng.* **23**, 10–21 (2015).

21. Seeber, M., Scherer, R., Wagner, J., Solis-Escalante, T. & Müller-Putz, G. R. EEG beta suppression and low gamma modulation are different elements of human upright walking. *Front. Hum. Neurosci.* **8**, 485–485 (2014).

22. Donati, A. R. C. et al. Long-term training with a brain-machine interface-based gait protocol induces partial neurological recovery in paraplegic patients. *Sci. Rep.* **6**, 30383 (2016).

23. McCrimmon, C. M. et al. Electrocorticographic encoding of human gait in the leg primary motor cortex. *Cereb. Cortex* **28**, 2752–2762 (2017).

24. Selfslagh, A. et al. Non-invasive, brain-controlled functional electrical stimulation for locomotion rehabilitation in individuals with paraplegia. *Sci. Rep.* **9**, 6782 (2019).

25. Greiner, N. et al. Recruitment of upper-limb motoneurons with epidural electrical stimulation of the cervical spinal cord. *Nat. Commun.* **12**, 435 (2021).

26. Moly, A. et al. An adaptive closed-loop ECoG decoder for long-term and stable bimanual control of an exoskeleton by a tetraplegic. *J. Neural Eng.* **19**, 026021 (2022).

27. Cappellini, G., Ivanenko, Y. P., Dominici, N., Poppele, R. E. & Lacquaniti, F. Migration of motor pool activity in the spinal cord reflects body mechanics in human locomotion. *J. Neurophysiol.* **104**, 3064–3073 (2010).

28. Daly, J. J. et al. Development and testing of the Gait Assessment and Intervention Tool (G.A.I.T.): a measure of coordinated gait components. *J. Neurosci. Methods* **178**, 334–339 (2009).

29. Larzabal, C. et al. Long-term stability of the chronic epidural wireless recorder WIMAGINE in tetraplegic patients. *J. Neural Eng.* **18**, 056026

(2021).

30. Harkema, S. et al. Effect of epidural stimulation of the lumbosacral spinal cord on voluntary movement, standing, and assisted stepping after motor complete paraplegia: a case study. *Lancet* **377**, 1938–1947 (2011).

31. Rejc, E., Angeli, C. A., Atkinson, D. & Harkema, S. J. Motor recovery after activity-based training with spinal cord epidural stimulation in a chronic motor complete paraplegic. *Sci. Rep.* **7**, 13476 (2017).

32. Angeli, C. A. et al. Recovery of over-ground walking after chronic motor complete spinal cord injury. *N. Engl. J. Med.* **379**, 1244–1250 (2018).

33. Gill, M. L. et al. Neuromodulation of lumbosacral spinal networks enables independent stepping after complete paraplegia. *Nat. Med.* **24**, 1677–1682 (2018).

34. Darrow, D. et al. Epidural spinal cord stimulation facilitates immediate restoration of dormant motor and autonomic supraspinal pathways after chronic neurologically complete spinal cord injury. *J. Neurotrauma* **36**, 2325–2336 (2019).

35. Shokur, S. et al. Training with brain–machine interfaces, visuo-tactile feedback and assisted locomotion improves sensorimotor, visceral, and psychological signs in chronic paraplegic patients. *PLoS ONE* **13**, e0206464 (2018).

36. Biasiucci, A. et al. Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. *Nat. Commun.* **9**, 2421 (2018).

37. Jovanovic, L. I. et al. Restoration of upper-limb function after chronic severe hemiplegia: a case report on the feasibility of a brain-computer interface controlled functional electrical stimulation therapy. *Am. J. Phys. Med. Rehabil.* **99**, e35–e40 (2020).

38. Nishimura, Y., Perlmutter, S. I., Eaton, R. W. & Fetz, E. E. Spike-timing-dependent plasticity in primate corticospinal connections induced during free behavior. *Neuron* **80**, 1301–1309 (2013).

39. Moorjani, S., Walvekar, S., Fetz, E. E. & Perlmutter, S. I. Movement-dependent electrical stimulation for volitional strengthening of cortical connections in behaving monkeys. *Proc. Natl Acad. Sci. USA* **119**, e2116321119 (2022).

40. Guggenmos, D. J. et al. Restoration of function after brain damage using a neural prosthesis. *Proc. Natl Acad. Sci. USA* **110**, 21177–21182 (2013).

41. Noga, B. R. & Guest, J. D. Combined neuromodulatory approaches in the central nervous system for treatment of spinal cord injury. *Curr. Opin. Neurol.* **34**, 804–811 (2021).

42. Barra, B. et al. Epidural electrical stimulation of the cervical dorsal roots restores voluntary upper limb control in paralyzed monkeys. *Nat. Neurosci.* **25**, 924–934 (2022).

43. Powell, M. P. et al. Epidural stimulation of the cervical spinal cord for post-stroke upper-limb paresis. *Nat. Med.* https://doi.org/10.1038/s41591-022-02202-6 (2023).

44. Knösche, T. R. & Haueisen, J. *EEG/MEG Source Reconstruction: Textbook for Electro- and Magnetoencephalography* (Springer, 2022).

45. Gramfort, A. et al. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, 267 (2013).

46. Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D. & Leahy, R. M. Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* **2011**, 879716 (2011).

47. Auboiroux, V. et al. Space–time–frequency multi-sensor analysis for motor cortex localization using magnetoencephalography. *Sensors* **20**, 2706 (2020).

48. Robinet, S. et al. A low-power 0.7 μV_{rms} 32-channel mixed-signal circuit for ECoG recordings. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **1**, 451–460 (2011).

49. Eliseyev, A. et al. Recursive exponentially weighted N-way partial least squares regression with recursive-validation of hyper-parameters in brain-computer interface applications. *Sci. Rep.* **7**, 16281–16281 (2017).

50. Sauter-Starace, F. et al. Long-term sheep implantation of WIMAGINE®, a wireless 64-channel electrocorticogram recorder. *Front. Neurosci.* **13**, 847 (2019).

51. Perry, J., Garrett, M., Gronley, J. K. & Mulroy, S. J. Classification of walking handicap in the stroke population. *Stroke* **26**, 982–989 (1995).

52. Tinetti, M. E., Franklin Williams, T. & Mayewski, R. Fall risk index for elderly patients based on number of chronic disabilities. *Am. J. Med.* **80**, 429–434 (1986).

53. Ermel, J. Von Fuß bis Kopf: Ganganalyse Teil 2: Anwendung in der Praxis. *physiopraxis* **4**, 30–34 (2006).

Acknowledgements

We thank our test pilot for his commitment and trust. Supported by Defitech Foundation, Rolex Award for Enterprise, International Foundation for Research in Paraplegia, Translational Medical Research Award 2021 from the Leenaards Foundation, Pictet Group Charitable Foundation, ONWARD medical, Medtronic, the Swiss National Science Foundation through the National Centre of Competence in Research in Robotics (51NF40-185543), Sinergia (CRSII5-183519), the Lead Agency Program with the French National Research Agency (Think2Move SNF-32003BE-205563, ANR-21-CE19-0038), A F Harvey Prize award, Swiss Innovation Agency InnoSuisse (CTI-41871.1 IP-LS Bridge), Eurostars (E!12743 Confirm and E!113969 Prep2Go), the European Commission (ERC-2019-PoC Braingait 875660, EIC 2021-TransitionChallenges-01-01 ReverseParalysis 101057450,

Horizon-EIC-2021-Pathfinderchallenges-01-02 NEMO-BMI 101070891), Fonds de dotation Clinatec (WIMAGINE implant development) and Institut Carnot Leti. ETH domain PHRT grant 2022-279.

Funding

Open access funding provided by EPFL Lausanne.

Author information

Author notes

1. These authors contributed equally: Henri Lorach, Andrea Galvez
2. These authors jointly supervised this work: Guillaume Charvet, Jocelyne Bloch, Grégoire Courtine

Authors and Affiliations

1. NeuroX Institute, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Geneva, Switzerland

   Henri Lorach, Andrea Galvez, Valeria Spagnolo, Nadine Intering, Molywan Vat, Cathal Harte, Salif Komi, Jimmy Ravier, Thibault Collin, Laure Coquoz, Icare Sakr, Edeny Baaklini, Sergio Daniel Hernandez-Charpak, Gregory Dumont, Leonie Asboth, Robin Demesmaeker, Jocelyne Bloch & Grégoire Courtine

2. Department of Clinical Neuroscience, Lausanne University Hospital (CHUV) and University of Lausanne (UNIL), Lausanne, Switzerland

   Henri Lorach, Andrea Galvez, Valeria Spagnolo, Nadine Intering, Molywan Vat, Cathal Harte, Salif Komi, Jimmy Ravier, Thibault Collin, Laure Coquoz, Icare Sakr, Edeny Baaklini, Sergio Daniel Hernandez-Charpak, Gregory Dumont, Leonie Asboth, Stefano Carda, Robin Demesmaeker, Jocelyne Bloch & Grégoire Courtine

3. NeuroRestore, Defitech Center for Interventional Neurotherapies, EPFL/CHUV/UNIL, Lausanne, Switzerland

   Henri Lorach, Andrea Galvez, Valeria Spagnolo, Nadine Intering, Molywan Vat, Cathal Harte, Salif Komi, Jimmy Ravier, Thibault Collin, Laure Coquoz, Icare Sakr, Edeny Baaklini, Sergio Daniel Hernandez-Charpak, Gregory Dumont, Leonie Asboth, Robin Demesmaeker, Jocelyne Bloch & Grégoire Courtine

4. Univ. Grenoble Alpes, CEA, LETI, Clinatec, Grenoble, France

   Felix Martel, Serpil Karakas, Olivier Faivre, Lucas Struber, Fabien Sauter-Starace, Vincent Auboiroux, Stephan Chabardes, Tetiana Aksenova & Guillaume Charvet

5. Medtronic, Minneapolis, MN, USA

   Rik Buschman, Nicholas Buse & Tim Denison

6. Department of Engineering Science, University of Oxford, Oxford, UK

   Tim Denison

7. Department of Rehabilitation, Sint Maartenskliniek, Nijmegen, the Netherlands

   Ilse van Nes

8. ONWARD Medical, Lausanne, Switzerland

   Anne Watrin

9. Univ. Grenoble Alpes, CHU Grenoble Alpes, Clinatec, Grenoble, France

   Lilia Langar & Stephan Chabardes

Contributions

H.L., A.G., V.S., M.V., N.I., T.C., L.C., I.S., E.B., S.D.H.-C., G.D., I.v.N., F.S.-S., V.A., S.Carda, J.B. and G.C. performed experiments and analysed data. H.L., A.G., F.M., S.Karakas, O.F., C.H., S.Komi, R.B., N.B., T.D., F.S.-S., V.A., T.A., R.D. and G.Charvet designed, developed and/or fabricated hardware and/or software. J.R., S.D.H.-C. and G.D. performed simulations. N.I., E.B., L.A. and I.v.N. supervised and conducted physical therapy. H.L., A.G., M.V. and A.W. managed regulatory affairs. H.L., A.G. and J.R. prepared illustrations. S.Chabardes and J.B. performed neurosurgical interventions. H.L., G.Charvet, J.B. and G.Courtine conceived and supervised the study. G.Courtine wrote the paper with J.B., H.L. and A.G., and all the authors contributed to its editing.

Corresponding authors

Correspondence to Guillaume Charvet, Jocelyne Bloch or Grégoire Courtine.

Ethics declarations

Competing interests

G.Courtine, J.B., H.L., R.D., L.A., T.A., F.M., G.Charvet and F.S.-S. hold various patents or applications in relation to the present work (EP4108289A1, EP2623025A1, EP2649936B1, EP3190480B1 and EP4088659A1). G.Courtine and J.B. are consultants for ONWARD medical. A.W. is an employee of ONWARD medical. G.Courtine and J.B. are minority shareholders of ONWARD, a company with potential commercial interest in the presented work. The remaining authors declare no competing interests.

Peer review information

Peer review information

Nature thanks James Guest, Nick Ramsey and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Technological and computational design underlying the *BSI*.

a, Photographs reporting the geometry and features of the WIMAGINE implant, including 64 platinum-iridium (90:10) electrodes with 4 mm x 4.5 mm pitch (in antero-posterior and medio-lateral axes respectively). Two external antennas are embedded within the implant. The first antenna powers the implanted electronics through inductive coupling at high frequency (HF, 13.56 MHz) while the second ultrahigh frequency antenna (UHF, 402-405 MHz) transfers the recorded signals outside the body. **b**, Two external antennas embedded in a personalized 3D-printed headset power the implant and recover the streamed signals that are then transferred to a base station. This base station manages the powering of the implants, synchronization and conditioning of the raw data. **c**, A decoding pipeline computes temporal, spectral and spatial features embedded in the ECoG signals related to the intention to move. These features are then uploaded into the decoding algorithm that decodes the attempts to move the lower limbs based on a tailored, recursive exponentially weighted Markov-switching multi-linear model algorithm[26]. This algorithm is a mixture of multilinear experts' algorithm integrating a Hidden Markov Model (HMM) classifier, called ***gating***, and a set of independent regression models, called ***experts***. The ***gating*** classifier predicts the joint that is intended to be mobilized (i.e. hip, knee or ankle on each side) as well as resting state, while each ***expert*** is dedicated to predicting the direction and relative amplitude of the intended movement. When updating is allowed, every 15 s, the coefficients of both linear regressions (β^{gate}, b^{gate}, β^{expert}, b^{expert}) are updated through recursive partial least square along with the coefficients of the transition matrix T corresponding to the number of transitions between each states during this 15s period (i.e. 150 new transitions). To support the production of standing

and walking, the outputs of the model are encoded into updates of joint-specific stimulation programs that are constrained within pre-established functional ranges of amplitudes. **d**, A tailored, medical-grade software sends these updates to the implanted pulse generator through a chain of wireless communication systems, eventually delivering the stimulation through a paddle array implanted epidurally over the lumbosacral spinal cord.

Extended Data Fig. 2 Calibration of the *BSI*.

a, Post-operative localization of the cortical implants over the segmented brain of the participant that confirms the appropriate positioning of the 64-electrode grids over the activated regions of the primary motor cortex responding to attempted lower limb movements, as measured during functional magnetoencephalographic recordings. **b**, Post-operative localization of the paddle lead over the lumbosacral spinal cord to target lower limb muscles. **c**, Projection of linear regression weights associated with different lower limb movements (depicted on body schemes) onto the location of the implants, revealing the spatial segregation of movement-specific features. **d**, Electromyographic activity recorded from several lower limb muscles following a burst of epidural electrical stimulation using the more selective electrode configurations (schemes) and parameters (reported) translated into polar plots reporting the amplitude of muscle responses. **e**, Spatial distribution of linear regression weights associated with upper versus lower limb movements over the grid of 64 electrodes from each cortical implant. The firmware enabled the selection of 32 electrodes within the 64 electrodes of each implant. The red dots indicate the 32 selected electrodes from each implant based on the amount of identified movement-related information for each of the 64 electrodes. **f**, Spectral distribution of linear regression weights associated with upper versus lower limb movements, highlighting the importance of high sampling density in low frequencies compared to high frequencies. This ensemble of features guided the parameterization of the decoders. **g**, Detailed representation of the spatial and spectral repartition of weights associated with decoding of the 6 different lower limb joint movements.

Extended Data Fig. 3 Stability of the decoder enables safe utilization of *BSI*.

a, Chronophotography and associated spectrogram, probabilities of left and right steps, modulation of muscle activity, ankle height, and peak probability of step cycles during a sequence involving walking, a voluntary pause (30 s, instructed), and resuming walking. The absence of false positive detections illustrates the robustness of the *BSI*. **b**, The bar plot reports the peak probability of walk (active) versus idle state, together with the confusion matrices reporting the detected rest versus left and right swing states (n = 31 and n = 49 samples for idle and active states respectively, unpaired one-tailed t-test ***, P < 0.001). **c**, Photographs illustrating sit to stand capacities without and with the *BSI*, including bar plots reporting balance capacities (scores) measured using the Berg Balance Scale.

Extended Data Fig. 4 The *BSI* normalizes gait parameters and supports walking on complex terrains.

a, Principal component (PC) analysis applied on kinematic and muscle activity parameters during walking on a treadmill with stimulation alone versus *BSI*. During stimulation alone conditions, a closed-loop controller based on motion sensors attached to the lower limbs determine the parameters of stimulation. Each dot represents a gait cycle. The bar plot reports the Euclidean distance in the PC space between each sample and the centroid of the healthy steps. (n = 119, n = 30 and n = 61 steps for healthy, EES only and *BSI* respectively, unpaired one-tailed t-test ***, P < 0.001). Compared to stimulation alone, the *BSI* enabled walking with gait features that were closer to those quantified in healthy individuals. This similitude is highlighted in the bar plots, which report the mean values of kinematic parameters with a high factor loading on PC1. **b**, Quantitative measure of step length while walking with crutches. Steps below 10 cm are considered failed as illustrated in the stick plot diagram. EES only condition showed significantly shorter step length due to increase of failed steps (n = 26, n = 43 for EES only and *BSI* respectively, Mann-Whitney U test one-tailed t-test **, P < 0.01). **c**, Photographs illustrating walking capacities, together with bar plots that report quantifications of performance during the execution of various walking paradigms, including walking up and down a ramp, climbing stairs, and walking with crutches overground. **d)** Walking on changing terrains with obstacles and different textures (6 surfaces), as illustrated in the scheme on the left. Conventions are the same as in previous

figures. Decoding stability is shown by overlayed probability curves of right hip flexions over consecutive steps (n = 13 steps, Left accuracy = 0.89 +/− 0.1 std, w = 2.06 s +/− 0.6 s std), and Left accuracy (n = 13 steps, accuracy = 0.91 +/− 0.1 std, w = 2.06 s +/− 0.4 s std).

Extended Data Fig. 5 Long term stability of the *BSI* supporting walking.

a, Recordings of the resting state were acquired regularly to evaluate the evolution of signal quality over time. Raw traces and power spectrum of an ECoG signal measured from a selected electrode are shown to illustrate the stability of the recorded signals. The plot reports the mean values of the power spectrum quantified over 2 min of resting state recorded at regular intervals over a period of nearly one year, showing a steady yet negligible decrease in signal quality over time (−0.03 dB/day). **b**, Plots reporting principal component analysis of gating coefficients from all the models used for supporting walking over the entire duration of the study. The size of each data point captures the relative time during which each model was used. **c**, Plots reporting the range of stimulation amplitudes and frequencies used over the entire course of the neurorehabilitation program, highlighting the robustness of the *BSI* over nearly six months of use. **d**, Spectrograms and decoding performance together with modulation of stimulation amplitude (relative) during self-paced walking enabled by the *BSI*. Plots report the probability of left and right hip flexion events (swing) measured over consecutive steps, and repeated at regular intervals over the entire time course of the clinical trial. **e**, Median spectrograms around the right hip flexion attempts during different time periods along training (n = 100 attempts in each period). The average rectified modulations show a significant increase with time (n = 64 electrodes, $R2 = 0.68$, $P < 0.001$).

Extended Data Fig. 6 *BSI* supporting control over isolated lower limb movements.

The same model was used to enable the participant to exert control over 6 joints from both sides during two sessions apart from 2 months. Conventions are the same as in previous figures.

Extended Data Fig. 7 Design and configuration of the *BSI* for independent use at home.

a, An integrated walker was designed and fabricated to incorporate the different hardware composing the *BSI*, thereby maximizing the practicability of the technological platform for use at home. The system is battery-powered and can operate autonomously for approximately 2 h without any supervision. **b**, Sequence showing the different steps to configure the *BSI*, including the positioning of the communication headset, uploading of a *BSI* program, monitoring of signal quality to ensure appropriate placement of the antennas, and adjusting the minimum and maximum amplitudes of the stimulation. The participant has been using the *BSI* independently to support neurorehabilitation and daily life activities over nearly one year. Positioning the hardware and configuring the *BSI* require approximately 5 min. **c**, Usage log and performance quantification of the participant after the main phase of the study as a cumulative number of decoded steps and cumulative time of use over a period of 181 day, i.e. since the participant returned to his home.

Extended Data Table 1 WIMAGINE device specifications
Extended Data Table 2 Epidural electrical stimulation system specifications
Extended Data Table 3 Observational gait analysis
Extended Data Table 4 Psychosocial Impact of Assistive Devices Scale (PIADS) questionnaire during the home-use phase

Supplementary information

Reporting Summary

Supplementary Video 1

Design of the brain–spine interface.

Supplementary Video 2

Implementation of the brain–spine interface.

Supplementary Video 3

Neurorehabilitation and neurological recovery enabled by the brain–spine interface.

Supplementary Video 4

Independent use of the BSI by the participant.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Lorach, H., Galvez, A., Spagnolo, V. *et al.* Walking naturally after spinal cord injury using a brain–spine interface. *Nature* **618**, 126–133 (2023). https://doi.org/10.1038/s41586-023-06094-5

- Received: 01 August 2022

- Accepted: 17 April 2023
- Published: 24 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06094-5

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

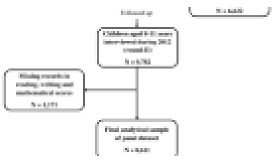
- **Brain–spine interface allows paralysed man to walk using his thoughts**
  - Dyani Lewis

  Nature (2023)

A digital bridge between brain and spinal cord restores walking after paralysis

Nature Clinical Briefing 24 May 2023

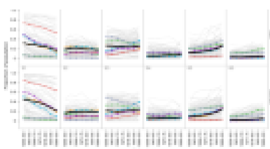
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06094-5



Early childhood circumstances and educational wellbeing inequality among tribal and non-tribal children in India: evidence from a panel study

14 June 2022

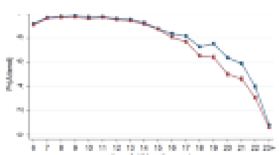
Rashmi Rashmi & Ronak Paul



Gender, education expansion and intergenerational educational mobility around the world

08 March 2023

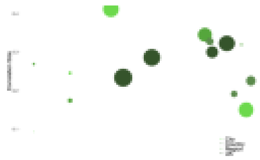
Yang Hu & Yue Qian



Predictors of the gender gap in household educational spending among school and college-going children in India

22 September 2022

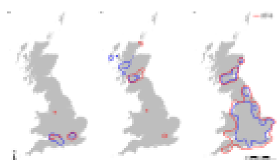
Rashmi Rashmi, Bijay Kumar Malik, … S. V. Subramanian



[Persistent association between family socioeconomic status and primary school performance in Britain over 95 years]

20 April 2022

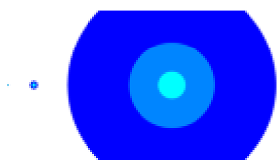
Sophie von Stumm, Sophie Nicole Cave & Paul Wakeling



[The geography of intergenerational social mobility in Britain]

26 October 2021

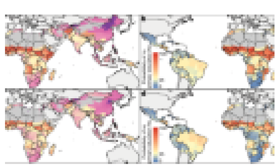
Paul A. Longley, Justin van Dijk & Tian Lan



[Dreams and realities of school tracking and vocational education]

28 February 2020

Mahmut Ozer & Matjaž Perc



Mapping disparities in education across low- and middle-income countries

25 December 2019

Local Burden of Disease Educational Attainment Collaborators

Does religion always help the poor? Variations in religion and social class in the west and societies in the global south

26 June 2018

Megan Rogers & Mary Ellen Konieczny

Addressing socioeconomic inequality in access to university education: an analysis of synergies and tensions in Scottish policy

18 December 2018

Edward M. Sosu, Lauren N. Smith, … Stephanie McKendry

- Article
- Open Access
- Published: 17 May 2023

Religion and educational mobility in Africa

- Alberto Alesina,
- Sebastian Hohmann,
- Stelios Michalopoulos &
- …
- Elias Papaioannou

Show authors

Nature **volume 618**, pages 134–143 (2023)

- 5141 Accesses
- 1 Citations
- 92 Altmetric
- Metrics details

Abstract

The African people and leaders[1,2] have long seen education as a driving force of development and liberation, a view shared by international institutions[3,4], as schooling has large economic and non-economic returns, particularly in low-income settings[5]. In this study, we examine the educational progress across faiths throughout postcolonial Africa, home to some of the world's largest Christian and Muslim communities. We construct comprehensive religion-specific measures of intergenerational mobility in education using census data from 2,286 districts in 21 countries and document the following. First, Christians have better mobility outcomes than Traditionalists and Muslims. Second, differences in intergenerational mobility between Christians and Muslims persist among those residing in the same district, in households with comparable economic and family backgrounds. Third, although Muslims benefit as much as Christians when they move early in life to high-mobility regions, they are less likely to do so. Their low internal mobility accentuates the educational deficit, as Muslims reside on average in areas that are less urbanized and more remote with limited infrastructure. Fourth, the Christian–Muslim gap is most prominent in areas with large Muslim communities, where the latter also register the lowest emigration rates. As African governments and international organizations invest heavily in educational programmes, our findings highlight the need to understand better the private and social returns to schooling across faiths in religiously segregated communities and to

carefully think about religious inequalities in the take-up of educational policies[6].

Main

Africa hosts vibrant Christian and Muslim communities and, given the demographic trends, it will be home to the largest numbers of both creeds in the coming decades[7]. Religiosity is high, with 75% of Christians and Muslims attending a church and a mosque at least once a week, according to the 2016 Afrobarometer Surveys. Nevertheless, the most important divide between African Muslims and Christians is not their religion. As we show here, educational gaps circumscribe the religious landscape. This is not a recent phenomenon. At independence, Christians enjoyed more schooling, reflecting colonial investments and missionary activity[8,9,10]. In several countries, mostly in West Africa, primary school completion for Christians was more than double that of Muslims or Africans adhering to local religions (Supplementary Information, section A). Although the share of Africans with no schooling has declined from two-thirds for those born in the 1950s to less than half for the 1990s cohort, religious differences persist. For example, in Nigeria (Africa's most populous country, roughly equally split between Christians and Muslims), the primary school completion rate of Christians is 0.88, whereas it is 0.57 for Muslims. In Ethiopia, Africa's second-most populous country, 29% of Christians (about two-thirds of the population) born in the 1990s have completed primary education, with Muslims registering 16%. African Muslims are more educated only in a few countries where they are small minorities—South Africa, Zambia and Rwanda (about 1–2% of the populations of these countries) and Uganda (11%).

Policy briefs and reports[7,11] have discussed the considerable Muslim–Christian differences in education levels but a comprehensive account of the evolution of interdenominational educational gaps and their determinants is missing. Although some pioneering case studies analyse the role of religion in Nigeria[12,13], looking among others at the Yoruba[14,15], there has been limited comparative work on religion, as the vibrant research in economics on the political economy of African development mainly focuses on ethnicity[16,17,18]. Similarly, research on the interplay between religion and

economic performance[19,20] often focuses on the role of Islam in the Middle East and Asia[21]. It is in Africa, though, that the interreligious gaps in education are striking and a plethora of narratives point to their rising salience[22]. An exception is the parallel work of refs. [23,24], who report considerable educational differences between Christians and Muslims in 11 African countries in 2017, showing further that gaps are considerable in predominantly Muslim areas. Focusing on Malawi, Nigeria and Uganda, these studies stress the role of religious leaders and social norms. Most importantly, the considerable body of research quantifying the role of educational reforms and government policies in the continent has not considered the role of religion[25]. Likewise, recent works mapping the intergenerational transmission of human capital across countries[26,27], US states[28], Chinese provinces[29] and African regions[30,31,32] do not study the role of religion. Others[33] explore regional, caste and religious differences in intergenerational mobility in education (IM; also referred to as educational mobility) across India; and like us, they uncover lagging educational mobility for the Muslim population, which contrasts with the rising mobility among the low socioeconomic status castes.

We construct statistics on faith-specific IM across 21 African countries and 2,286 regions and uncover substantial differences between Muslims, Christians and Africans adhering to Traditional (Folk) religions (Animists). (We use the terms Traditionalists and Animists interchangeably throughout). We then trace the roots of these disparities. Interreligious differences in education levels at independence explain a substantial fraction of the observed variation. Nevertheless, even when we compare young Africans living in the same district with similarly (un)educated elders in their religious group, in households of comparable size and structure and in which the household head works in the same broad sector and occupation, IM gaps, albeit attenuated, endure. Muslim children underperform in regions where they are numerous compared to where they are a minority, a pattern that is not present for Christians and adherents of Traditional religions. Muslims benefit as much as Christians when they move early in life to high-IM regions. Given the explanatory power of religious segregation and the impact of regions on adherents of all creeds, we conclude with a primer on internal migration differences. Christians are much more likely to emigrate and exploit opportunities outside their birthplace in almost all countries. The

low propensity of Muslims to move accentuates their initial educational disadvantage, as they typically reside in remote places, far from the capital and the coastline, with limited missionary activity and transportation investments. Muslims register the lowest emigration rates and largest IM deficits in these religiously segregated areas.

Data and educational mobility statistics

We construct measures of IM across 2,286 districts in 21 African countries using information from the Integrated Public Use Microdata Series (IPUMS) International, which collects and harmonizes censuses, reporting representative (typically 10%) samples (see 'Data availability' in Methods). As the detail of religious denomination differs across censuses, we aggregate into Muslims, adherents of Traditional religions, Christians and two auxiliary categories (Other and No Religion). Extended Data Table 1 and Extended Data Fig. 1 give the shares in our 21-country sample (see also Data section). Our measures of educational IM reflect how children fare vis-a-vis their cohabiting elder members, mainly parents, in the household. We use primary school completion as the critical educational milestone for Africans born up until the 1990s, as secondary and college enrolment has increased mainly since the 2000s (Supplementary Fig. A2). Upward IM denotes the likelihood that 14–18-year-old children whose parents have not completed primary education manage to complete primary school; downward IM gives the likelihood that 14–18-year-old individuals whose parents have completed primary education will not manage to finish primary school. Our base sample consists of 7,188,717 children between 14 and 18 years of age. We also look at 14–25-year-old children as the sample increases to 13,018,904 with cohabitation rates of about 75%. We provide details in the Data and educational IM measures section and in Supplementary Information, section B.

Religious IM across Africa

Country patterns

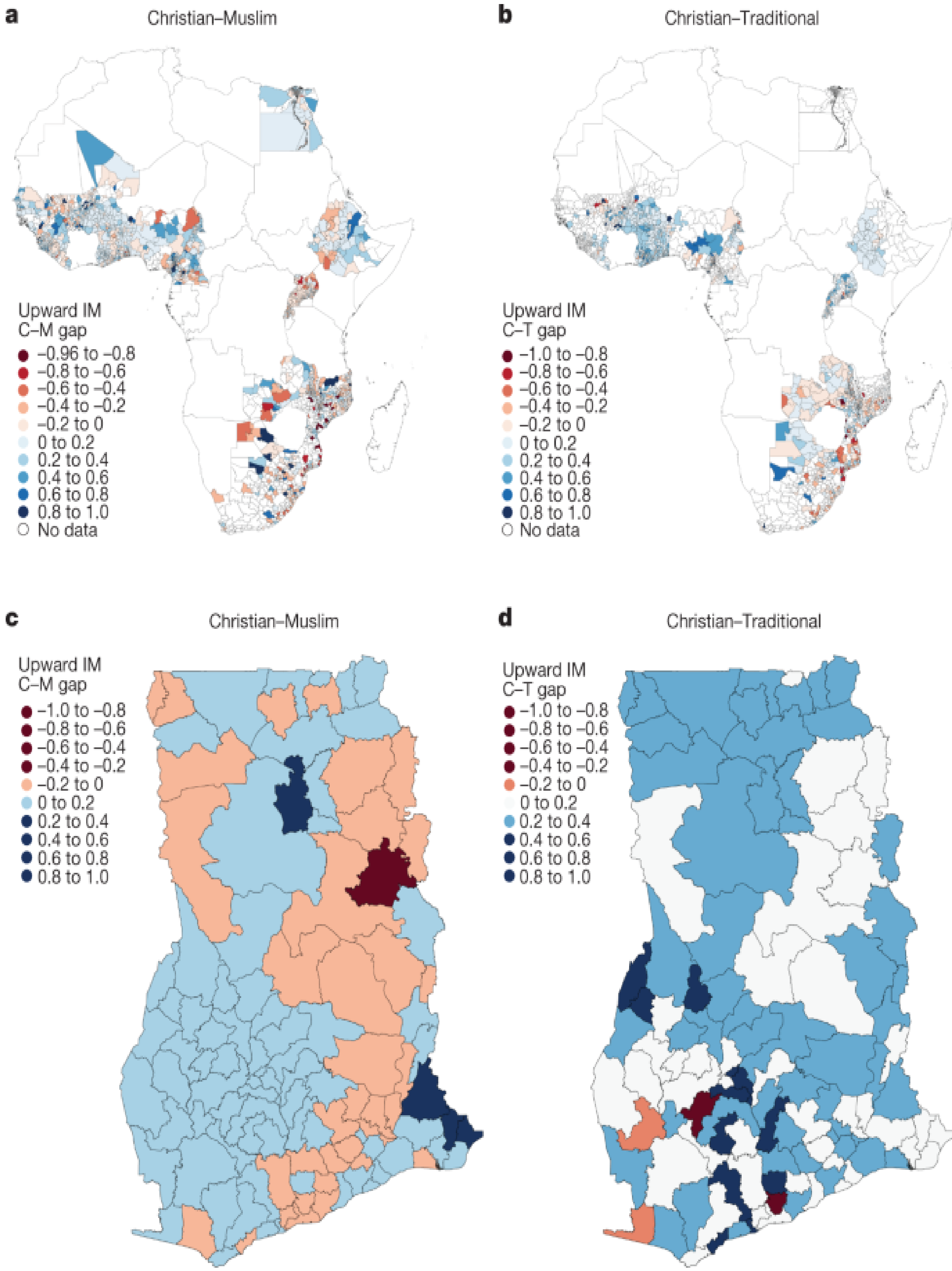
Table 1 reports the newly constructed faith-based IM statistics across countries, the Christian–Muslim gaps in upward and downward IM and their level of statistical significance (Supplementary Information, section C provides additional measures.) With the exception of Mozambique, upward IM is lowest for Africans adhering to Traditional religions (column 4 of Table 1); it is less than 10% in Burkina Faso, Sierra Leone, Rwanda, Malawi, Uganda and Ethiopia. Upward IM is low for Traditionalists, even in countries with considerable representation. For example, in Togo where overall IM is 52.6%, Animists register 38.2%; and in Benin their 21.3% IM trails the country average by 8.5 percentage points. Christians enjoy the highest upward IM in 15 out of the 21 countries with a cross-country mean of 41.2%. The IM of Muslims exceeds that of Christians in South Africa (87.4% versus 74%), Zambia (47.6% versus 44.5%, although this difference is not significant at standard confidence levels) and Rwanda (27.4% versus 18.3%), where their shares are tiny (0.8–1.5%) and in Uganda (48.5% versus 40.4%), Mauritius (96% versus 87.8%) and Liberia (26.6% versus 21.8%), where Muslims constitute between 10% and 16% of the population. The Christian–Muslim gap in upward IM is considerable in many countries. In Nigeria, Muslim children of illiterate parents are 32 percentage points less likely to complete primary schooling than Christian children born to similarly uneducated parents (78.6% versus 46.6%). In Ethiopia, a country with very low primary school completion rates, the upward IM for Christians is 13.8%, whereas for Muslims it is 8.2%. The Christian–Muslim upward IM gap is largest in West Africa (that is, the mean is 22.1%), where Muslims form the majority (Guinea, Senegal and Burkina Faso) or substantial minorities (Nigeria, Benin, Cameroon and Ghana). Downward IM is high for Muslims (the mean is 27.5%) and Traditionalists (the mean is 42.6%). In West Africa, roughly one out of six Muslim children born to parents who completed primary schooling will fail to do the same. In Cameroon, the downward IM is 4.1% for Christians and a staggering 19.6% for Muslims. Downward IM for Nigerian Muslims is twice that of Christians (16.2% versus 7.8%). This pattern is reversed in countries with small Muslim communities.

Table 1 Country-group-level estimates of IM, ages 14–18 yr

Regional patterns

Figure 1a portrays, for the 1980 and 1990 birth cohorts, the gap in upward IM between Christians and Muslims and Fig. 1b between Christians and adherents of Traditional religions across 1,731 and 1,071 African districts, respectively. (Supplementary Table C3 summarizes the districts' population and land area. Within countries, Muslim-majority districts are comparable in area to Christian). The depicted districts are fewer than the total number of districts as we need to observe in each district 14–18-year-old individuals born in the 1980 and 1990 cohorts from both religions. Figure 1c,d zoom into Ghana to illustrate the within-country variation. The Christian–Muslim gap is large in West Africa. (Supplementary Fig. C1 maps upward IM for Christians, Muslims and Traditionalists, respectively). Extended Data Table 2 reports regional statistics. Within-country variation is notable in Cameroon, Ghana, Guinea, Nigeria and Ethiopia. Christians fare better than Muslims in almost all the districts of Senegal and Burkina Faso, despite the numerical dominance of Muslims. In Ghana and Cameroon, the upward IM of Christians exceeds that of Muslims in three out of four districts. The Christian–Muslim upward IM gap is lower in the southern districts of West African countries, an issue we return to in the section on Religious IM across Africa. Educational outcomes for Muslims, Traditionalists and Christians move together as the cross-region correlation of religious upward IM is 0.60–0.72 and for downward IM 0.39–0.62 (Supplementary Table C6). However, Muslims and Traditionalists underperform compared to Christians, even when we zoom in on the same district. In the median region, upward IM for Christians is 0.44, 0.33 for Muslims and 0.21 for Traditionalists. Downward IM for Christians is 0.25, for Muslims 0.29 and 0.35 for Traditionalists (Supplementary Table C4).

Fig. 1: Religious IM gaps across African districts.



a,**b**, Differences in educational upward IM between Christians and Muslims (**a**) and Christians and Africans adhering to Traditional religions (**b**) across

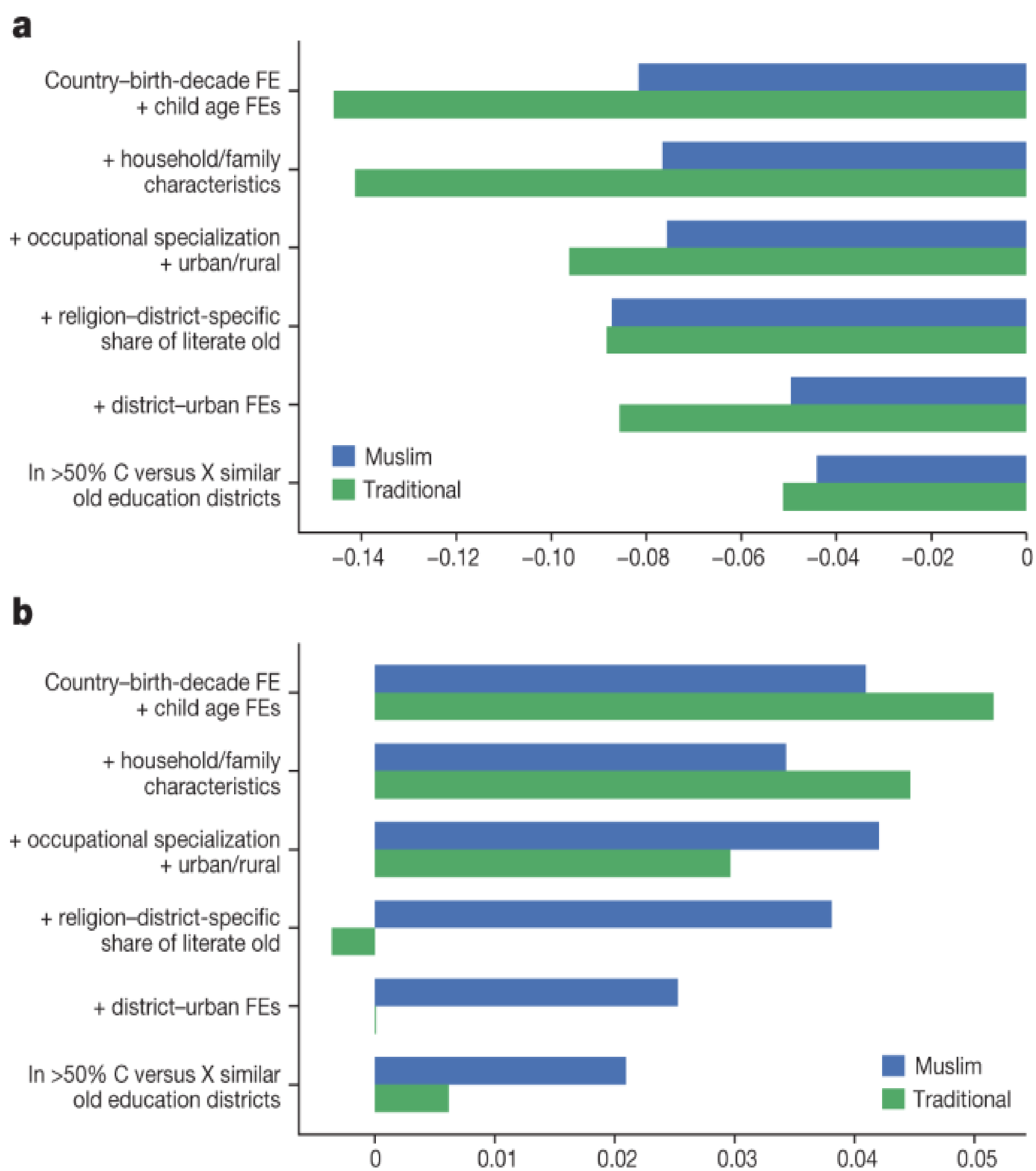
all African districts (administrative level-2 and level-3 regions) in 21 countries. **c**,**d**, The Christian–Muslim (**c**) and the Christian–Traditionalist (**d**) difference in upward IM across districts in Ghana. We estimate absolute upward IM for young individuals, aged 14–18 yr, cohabiting with at least one older generation member in the household, usually a parent. Supplementary Figure C1 provides the mappings of upward IM for Christians, Muslims and Traditionalists used to compile the maps of Christian–Muslim and Christian–Traditionalist gaps in educational mobility.

Explaining the religious IM gaps

Approach

We estimate individual-level regressions that associate the IM of children with their own religious affiliation, using Christians as the omitted category. We drop ‘other religion’ and ‘no religion’ as these are not comparable across censuses and countries. A large body of research documents non-negligible differences in family arrangements, age of marriage, social practices and occupational specialization between Christians, Muslims and adherents of Traditional religions (Animists). To explore their role, we examine how the estimates on the Muslim and the Traditional religion indicators change as we account for: (1) household size and family structure variables, including the age of marriage; (2) proxies of household income (as, for example, the profession and the industry of employment of the household head); (3) the literacy of the older generation for the individual’s coreligionists in the district; and (4) district constants interacted with urban–rural household status. These features may interact or be jointly determined by deeper factors. Our objective is not to identify causal effects but to examine the fraction of the variation in educational mobility gaps explained by these aspects. Figure 2 reports ordinary least-squares (OLS) estimates weighting by the population of a country to account for differences in IPUMS coverage, whereas Extended Data Fig. 2 reports country-specific estimates of Christian–Muslim differences. (Supplementary Information, section D gives results for various data subsets).

Fig. 2: Drivers of religious IM gaps.



a,**b**, Illustration of how the Christian–Muslim (blue bars) and the Christian–Animist (green bar) differences in upward IM (**a**) and downward IM (**b**) change as we add controls for household and family structure (model 2), parental occupational specialization, the industry of employment and rural–urban residence (model 3), the district's religion-specific illiteracy rate among the older generation (model 4), district times rural/urban fixed effects (model 5) and focusing on the top 50% most similar districts in each country in terms of the literacy gap of the elders in each religious group (model 6).

The bars on the top (model 1) reflect the baseline interreligious differences in IM, simply conditioning on country–birth cohort fixed effects (FE) and age constants. The sample consists of Muslim, Animist and Christian young individuals (14–18 yr), matched to the previous generation in the household. The figure gives weighted linear probability model (OLS) estimates using the 1980 populations of the countries for the weighting to account for differential IPUMS sampling/coverage across countries. In models 1–5, the upward IM regressions are run in a sample of 4,989,952 young individuals who cohabit with older generation members who have not completed primary education and the downward IM regressions in a sample of 1,919,711 young individuals cohabiting with older generation members who have completed primary education. The figures omit standard-error bands to enable clear visualization of the patterns. The estimates on the Muslim indicator are statistically significant across all 12 specifications, whereas the Animist coefficient is significant across all upward IM specifications. C, Christian; X, Muslim or Animist.

Baseline IM gaps

The top bars simply condition on country cohort fixed effects and children's age constants to account for the increase in education over time. Among children whose parents have not completed primary education, Christians enjoy an advantage of 16 percentage points in completing primary school over Muslims and 20 percentage points over adherents of Traditional religions born in the same decade and country. Muslim children whose parents have completed primary education are 7 percentage points less likely than Christians to achieve the same qualification. If anything, Traditionalists have a lower rate of downward mobility than Muslims.

Household structure and size

We condition on a rich set of household features: (1) size, distinguishing between individuals of the generation of the young, the head and the grandparents; (2) indicators for each 14–18-year-old individual regarding his/her relationship to the household head (for example, biological children, relative, nephews, spouse); (3) family organization indicators (for example, father and mother present, father or mother only, others only); and (4)

mother's and father's age at their offspring's birth. We include all variables concurrently, as our objective is not to identify the most relevant one but to examine how much the religion coefficients move when we account for a saturated set of household traits. As we show in Supplementary Information, section D, there are important interreligious differences in household features. Muslim and Animist households are, on average, larger; 14–18-year-old Muslim girls are more likely to be spouses of the household head than Christian girls of the same age; and Muslim and Animist parents are, on average, younger at their children's birth than for Christians, reflecting earlier marriages. Despite these differences in the family organization, the educational gaps do not move much when we account for the former. The upward IM gap drops from 0.165 to 0.15 for Muslims and from 0.20 to 0.18 for adherents of Traditional religions. Household structure and size play a somewhat bigger role in narrowing the Christian–Muslim IM gap in West Africa but their influence is negligible in East and Southern Africa (Extended Data Fig. 2). Downward IM differences between Muslims and Christians are more affected by the inclusion of the rich set of household controls closing the overall gap by roughly 20% from 0.07 to 0.055.

Household and parental occupational differences

We then turn to economic features, conditioning in model 3 of Fig. 2 on the following three (sets of) attributes (1) urban versus rural residence; (2) the industry of employment (six categories); and (3) the occupational specialization (ten categories) of the household head. As shown in Supplementary Information, section D, adherents of Traditional religions are more likely to live in rural areas (20%) and work in agriculture and are less likely to be professionals or skilled employees. Christian–Muslim differences in the employment industry, rural–urban status and profession are minor, perhaps masking larger ones owing to the coarse aggregation of IPUMS. In line with these patterns, differences between Christians and Muslims in living conditions and access to necessities in Afrobarometer Surveys and household income in the surveys[7] are also small (although Muslim households are larger). Variations in the industry of employment (broadly defined), occupation and urban–rural residence explain a non-negligible component of the IM differences between Christians and Traditionalists: the upward IM gap drops from 0.20 to 0.15

and the downward IM gap halves. These economic features do not explain the Christian–Muslim IM upward gap. If anything, the downward mobility gap becomes more pronounced when we compare children born in households of similar occupational structures.

Initial literacy

In model 4 of Fig. 2, we control for the share of the older generation with completed primary education in the district for each religious group. This accounts for initial, group-specific regional development and schooling, stemming from the location of Christian missionaries, colonial educational investments and the spread of Islam[34]. The Christian–Muslim upward IM gap drops from 0.15 to 0.085. Likewise, the Christian–Muslim downward IM gap goes from 0.07 to 0.04, whereas the Christian–Traditionalist downward IM is eliminated. Differences in religious literacy rates of the older generation across districts seem first order, consistent with our earlier work[31], showing that initial literacy is the most important correlate of regional IM.

Regional features

As recent studies[30,31,32] show that regions have a chief role in educational mobility, in model 5 of Fig. 2, we augment the specification with district-specific constants interacted with urban indicators (the interaction with the urban indicator does not affect the estimates). Regions explain a non-negligible fraction of the IM gaps, as there is considerable segregation across most African countries[35] and Muslims (and Traditionalists) reside in places with low opportunity (see below). The coefficient on the indicator for Traditional religions in the upward IM specification is not much affected (stays around 0.08), whereas the downward IM gap is eliminated. Accounting for regional features reduces the Christian–Muslim upward IM gap to 0.05 and the downward IM to 0.03. The explanatory power of regions for the Muslim–Christian mobility gaps seems first order in many countries, mainly in West Africa, where religious segregation is the highest (Extended Data Fig. 2). In Benin, Cameroon, Ghana and Senegal, the Christian–Muslim gap in upward IM halves when we account for residence (and initial differences in literacy). In Nigeria, the upward IM gap drops from 0.30 to

0.10, whereas in Mozambique, it goes from 0.08 to nil. This pattern echoes the patterns across India[36] of considerable heterogeneity in relative educational mobility across castes and religions in narrow geographic areas. In the last row, we restrict estimation to half of the regions where interreligious differences in completed primary education of the older generation are the smallest to better account for 'initial' conditions in the district. The Christian–Muslim gap in upward IM continues to be close to 5 percentage points and 2 percentage points for downward IM.

Taking stock

The analysis shows the following. First, differences in household structure between Muslims and Traditionalists, on the one hand, and Christians, on the other, explain a small fraction of educational mobility differences. Second, broad economic features have no role in the Christian–Muslim upward gap but reduce the Christian–Animist difference. Third, the literacy of the older generation in the district for each religious group and region-specific features explain roughly two-thirds of the interreligious IM gaps; this is especially the case in West Africa, where segregation is the highest. Fourth, Christian–Muslim educational gaps remain, even when we compare children with the same parental background and family composition residing in the same region (with comparable literacy rates in the older generation of the various religious groups).

Sorting and childhood regional exposure

The prominence of regions gives rise to a plethora of questions ranging from their independent impact on the educational IM of the different denominations to unpacking the relevant regional characteristics. In this section, we address the former and, in the next section, the latter. Supplementary Information, section E gives descriptive statistics and further evidence.

Empirical design

To isolate regional childhood exposure effects from spatial sorting, we use the approach of refs. [37,38] and leverage differences in the age of children's moves across districts[39]. The specification, detailed in the Methods, associates primary school completion for children of uneducated parents (upward educational IM) with the age of children's move and differences in educational mobility between birthplace (*o*) and destination district (*d*) among non-movers of the same cohort b $[{\varDelta }_{odb}={\overline{{\rm{IM}}}_{{\rm{up}}}}_{bd}^{{\rm{nm}}}-{\overline{{\rm{IM}}}_{{\rm{up}}}}_{bo}^{{\rm{nm}}}]$. ${\overline{{\rm{IM}}}_{{\rm{up}}}}_{b}^{{\rm{nm}}}$ summarizes all features of the economic, social, educational and institutional environment, shaping IM in a given district. These could reflect, among many others, local returns to schooling, school quality and quantity, accessibility, schooling fees and teacher–pupil ratios that differ considerably in low-income countries[40].The age-specific parameters on destination minus origin IM, ${\beta }_{m}^{{\rm{rel}}}$, capture how a child's probability of completing primary schooling varies with the age of their move to districts with higher or lower mobility. If regions matter for mobility, the earlier the move, the greater the impact. The variation comes from children born in the same place and decade moving to regions with different IM. Differences in the age-of-move slopes reflect the impact of an extra year in the high IM district—regional childhood exposure effects. The main identifying assumption, backed by our earlier work[31], is that the timing of moving for households is unrelated to children's ability.

Sample

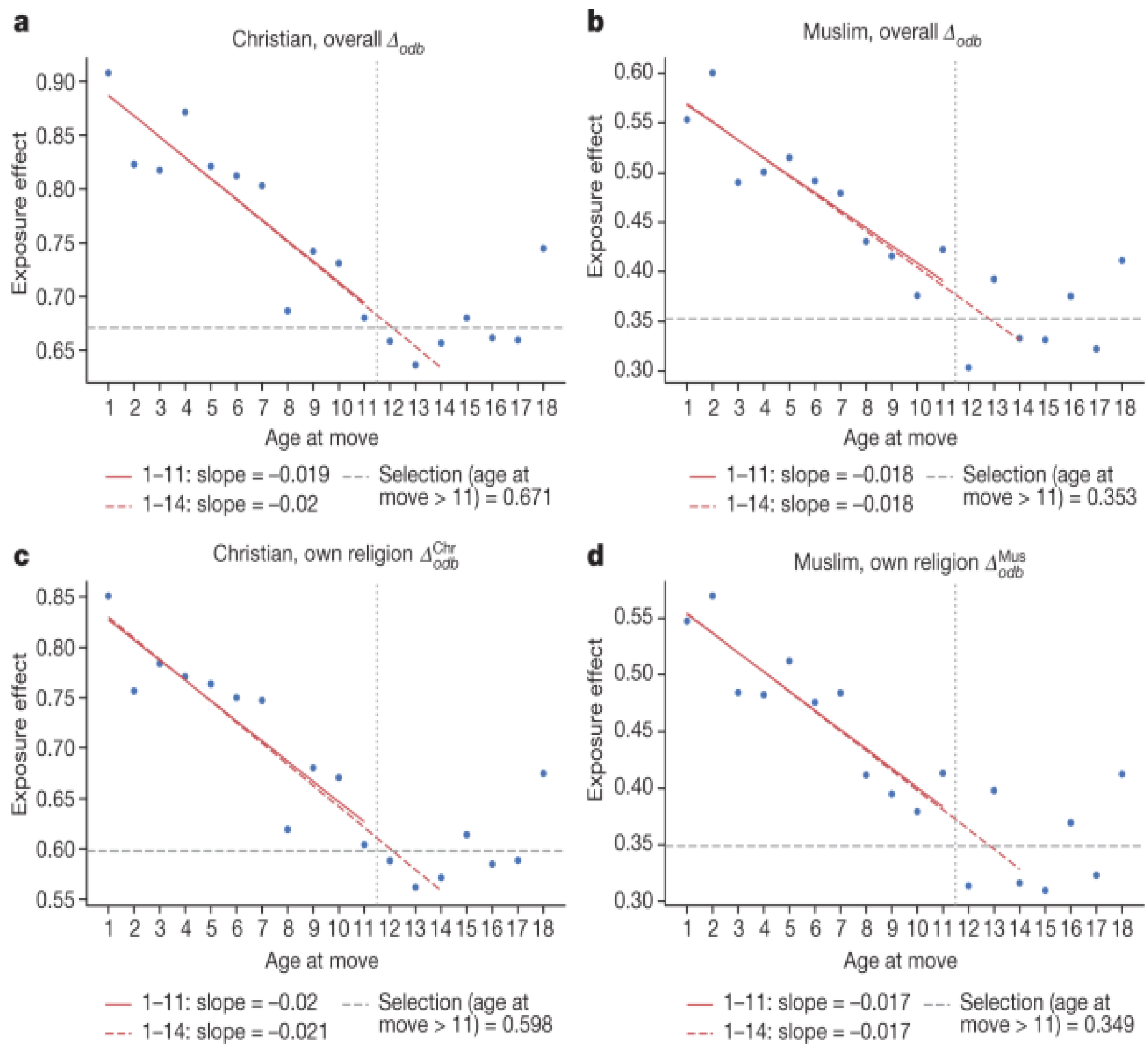
Not all censuses record the age of the move. These estimates come from a sample of 276,686 14–25-year-old individuals from 13 countries. These countries come from all major African regions; they are both former British and French colonies and non-colonized (Ethiopia); relatively poor and rich.

Results

Figure 3 plots the age-of-move estimates, ${\hat{\beta }}_{m}^{{\rm{rel}}}$, for Christians and Muslims against the child's age at the time of the household move. In Fig. 3a,b, origin–destination upward

IM differences are calculated using all non-movers independently of religious affiliation, whereas in Fig. 3c,d, we use non-movers of the same religious group. The estimates for Traditionalists (not reported) are similar but imprecise, reflecting the small sample and the blending of heterogeneous religions. The figure uncovers two sets of patterns. First, the age-of-move slopes, ${\hat{\beta }}_{m}^{{\rm{rel}}}$, are large for early-in-life moves and decline until around the age of 12–14 years. Children moving to 'better' districts earlier in life have a higher propensity to complete primary schooling. The relationship between age at move and exposure is negative and approximately linear, implying 'regional childhood exposure effects': moves in higher (lower) mobility districts are beneficial (detrimental) for younger kids. A further year of exposure before the age of 12 years for a child of illiterate parents to higher mobility district increases her chances for completing primary school by roughly 2 percentage points. Exposure effects are also 2% for Muslims and Christians when we compare them to their coreligionists. Both Muslim and Christian children benefit (lose) from early moves into high (low) upward mobility regions. Extrapolating over 14 years of childhood, Muslim and Christian children who move at birth to a district with one percentage point higher upward IM among Muslims will pick up roughly 30% of this difference through the impact of the region. Second, the slopes are significantly positive, even for children moving after 13–14 years. As moves after that age are unlikely to affect primary education, they reflect selection. Parametric specifications imposing a piecewise linear structure of exposure effects from 1 to 11 and from 12 to 18 yields similar regional childhood exposure estimates for Christians and Muslims. Besides, within-household specifications exploiting variation in the time of move across siblings also yield regional childhood exposure effects of about 2% for both faiths (Supplementary Information, section E).

Fig. 3: Regional childhood exposure effects, semiparametric estimates.



Estimates of childhood regional exposure effects of upward IM. **a**–**d**, The figures plot age-of-move slopes on differences in upward IM (blue circles), estimated among non-movers of all faiths in **a** and **b** and of the same religion in **c** and **d** between origin and destination district against the child's age when the household moved. The dependent variable takes the value of one when the child of a household in which the older generation has not completed primary education completes primary schooling (upward IM). Panels **a** and **c** give estimates for Christians and **b** and **d** for Muslims. The dashed vertical line in **a**–**d** separates the data into moves before the age of 12 yr that are most relevant for primary education and after the age of 12 yr. Also shown is the OLS regression fit of age-of-move slopes against the age of move before age 12 (red solid line) and age 14 (red dashed line). The slopes represent estimates of regional childhood exposure effects for primary

school completion by ages 12 and 14 yr. The age-of-move estimates are statistically significant at standard confidence levels. The panels omit standard errors to avoid cluttering. The results for Christian children are based on 138,300 and 141,355 individual observations in the specifications with overall differences in non-movers and Christian-only non-movers, respectively. The Muslim results are based on 127,914 and 128,215 observations in the specifications with differences in IM among all non-movers and differences in IM among Muslims, respectively.

Taking stock

We obtained two results: (1) although spatial sorting is considerable, children whose families move earlier to areas where residents (of all faiths or of their religion only) have higher IM are more likely to complete primary school and (2) regional childhood exposure effects are comparable for Christians and Muslims.

Religious IM gaps across African regions

Why do region-specific constants explain roughly half of the religious IM gaps? There are two possible explanations: (1) religious segregation, coupled with regional differences in educational opportunity, arising, for example, from more accessible and better-quality schools, make upward mobility more challenging for Muslims and adherents of Traditional religions and (2) the same regional feature may influence followers of different religions differently.

Differences in residence across faiths

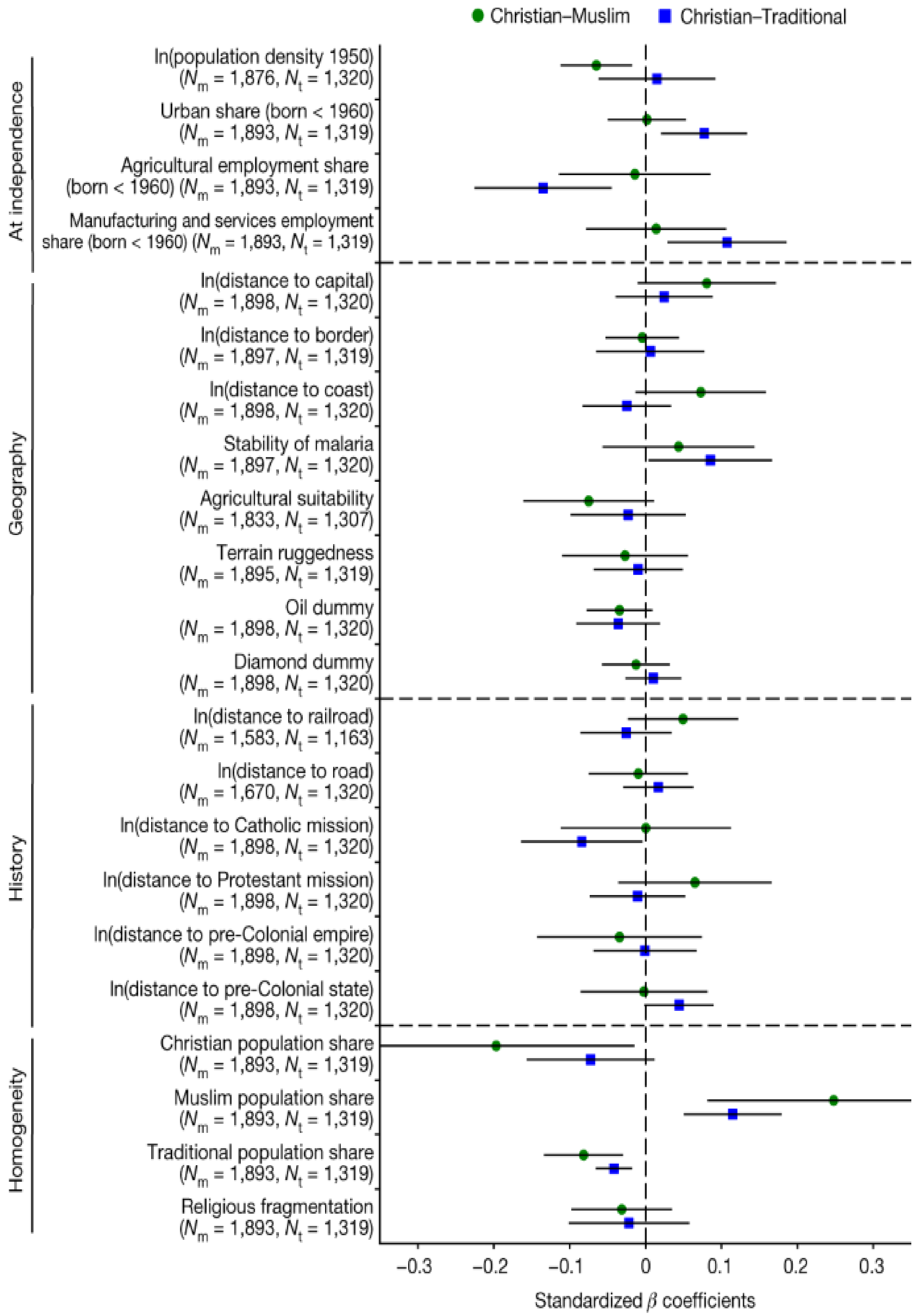
In Methods and in Extended Data Fig. 3 we explore how the residence characteristics compare across religious groups. On average, Muslims and Traditionalists reside in regions that, at the end of the colonial period, were less densely populated and more agriculture-oriented than Christians. Besides, Muslims and Traditionalists reside away from the capitals and the coastline; they live in districts with fewer colonial investments, away from colonial railroads, roads and Christian missions. Consequently, the lower

representation of Muslims in regions with better initial conditions is of first-order importance for explaining the observed Christian–Muslim IM gap.

Approach

We estimate within-country district-level specifications associating regional Christian–Muslim and Christian–Animist differences in upward IM to various regional characteristics. Figure 4 reports standardized coefficients. As the literacy rate of the older generation is the strongest correlate of IM (Fig. 2), we also run specifications conditioning on it (Extended Data Fig. 6). Extended Data Figures 4 and 5 present regional specifications separately for adherents of Christianity, Islam and Traditional religions and compare the magnitudes of the coefficients. Although the analysis does not have a causal interpretation, it allows for characterizing the geography of the considerable variation in religious educational IM within countries. Besides, it complements the evidence in Fig. 3, as the movers' design that distinguishes spatial sorting from childhood exposure effects does not pin down which regional features correlate with religious IM.

Fig. 4: District correlates of Christian–Muslim and Christian–Traditionalist upward IM gaps.



Christian–Muslim
Christian–Traditional
At independence
ln(population density 1950) (N_m = 1,876, N_t = 1,320)
Urban share (born < 1960) (N_m = 1,893, N_t = 1,319)
Agricultural employment share (born < 1960) (N_m = 1,893, N_t = 1,319)
Manufacturing and services employment share (born < 1960) (N_m = 1,893, N_t = 1,319)
Geography
ln(distance to capital) (N_m = 1,898, N_t = 1,320)
ln(distance to border) (N_m = 1,897, N_t = 1,319)
ln(distance to coast) (N_m = 1,898, N_t = 1,320)
Stability of malaria (N_m = 1,897, N_t = 1,320)
Agricultural suitability (N_m = 1,833, N_t = 1,307)
Terrain ruggedness (N_m = 1,895, N_t = 1,319)
Oil dummy (N_m = 1,898, N_t = 1,320)
Diamond dummy (N_m = 1,898, N_t = 1,320)
History
ln(distance to railroad) (N_m = 1,583, N_t = 1,163)
ln(distance to road) (N_m = 1,670, N_t = 1,320)
ln(distance to Catholic mission) (N_m = 1,898, N_t = 1,320)
ln(distance to Protestant mission) (N_m = 1,898, N_t = 1,320)
ln(distance to pre-Colonial empire) (N_m = 1,898, N_t = 1,320)
ln(distance to pre-Colonial state) (N_m = 1,898, N_t = 1,320)
Homogeneity
Christian population share (N_m = 1,893, N_t = 1,319)
Muslim population share (N_m = 1,893, N_t = 1,319)
Traditional population share (N_m = 1,893, N_t = 1,319)
Religious fragmentation (N_m = 1,893, N_t = 1,319)
−0.3 −0.2 −0.1 0 0.1 0.2 0.3
Standardized β coefficients

Correlations (standardized β coefficients) between upward IM gaps of Christians and Muslims (green circles) and Christians and Traditionalists (blue squares), averaged across individuals in a district and four sets of regional characteristics: (1) proxies of development at independence (for example, population density and urban share); (2) location and geographic features (for example, distance to coast, border and capital); (3) historical aspects reflecting colonial-era investments and precolonial statehood; (4) religious homogeneity, proxied by the share of each religious group in the district and a frationalization index. Supplementary Information, section F gives variable definitions and sources. All specifications include country fixed effects (constants not reported). The 2 s.e. bands based on heteroskedasticity adjusted clustered at the country level are also reported. The estimates (green and blue symbols in the figure) were obtained by running separate regressions of the district-level Christian–Muslim and Christian–Traditionalist IM gap, respectively, on each district-level variable (indicated on the vertical axis of the figure). The IM gap is defined as the average IM of Christians minus the average IM of Muslims or Traditionalists in the district. Before running each regression, we standardize the dependent and independent variable by subtracting its sample mean and dividing it by its sample standard deviation.

Early development, historical investments and geography

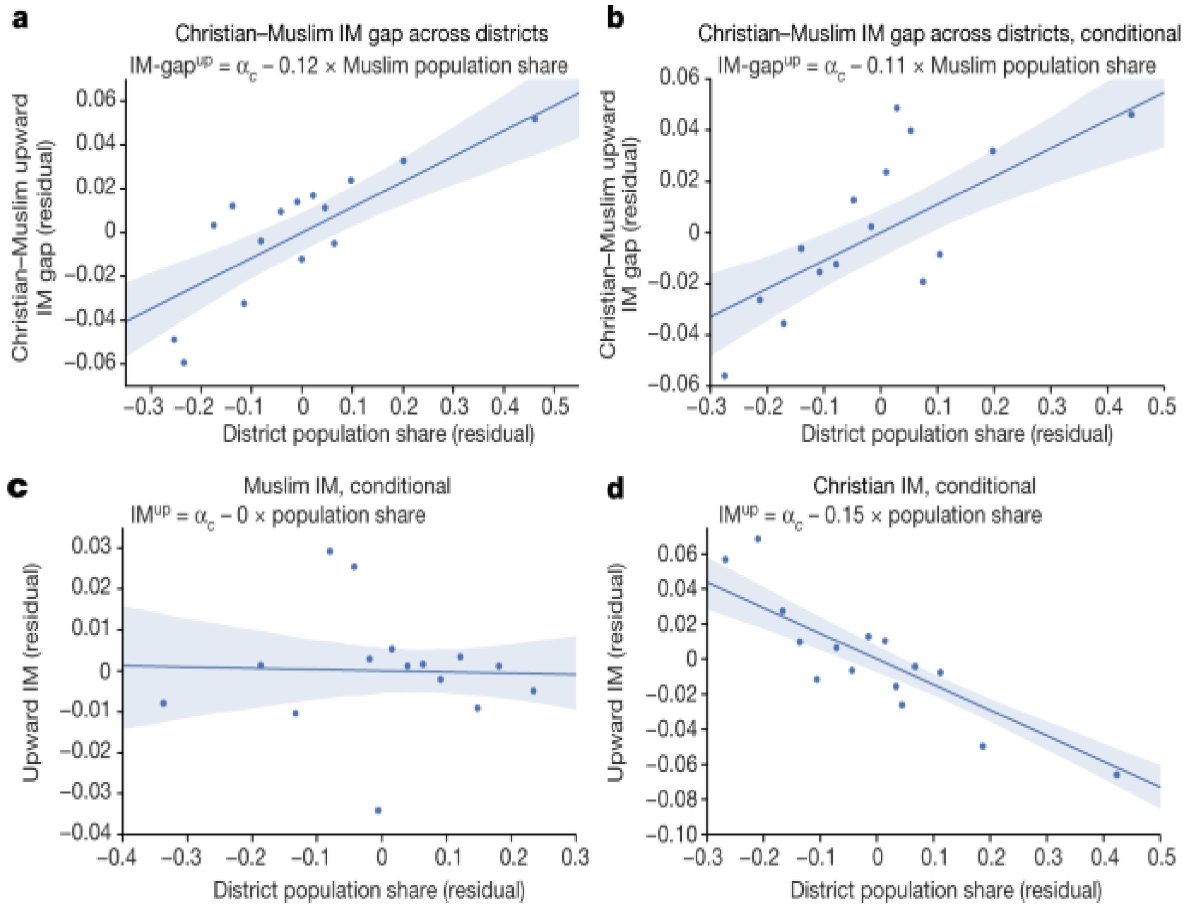
Drawing on research on the roots of African development[41], we start examining the role of geographic, historical and at-independence features. As Extended Data Fig. 4 shows, upward IM is higher and downward IM is lower in more densely populated and urbanized regions, more specialized in services and manufacturing, close to the capitals, the coast and missionary activity and transportation infrastructure. However, these features do not explain the Christian–Muslim and the Christian–Traditionalist IM gaps, as the correlations are similar across groups.

Segregation

We then turn to religious fragmentation and segregation. Our exploration is motivated by three observations. First, religious IM gaps are larger in segregated countries and Muslim educational mobility is the highest in

countries with small Muslim communities. Second, US-centred research shows that racial segregation moves in tandem with underinvestments in education[42,43]. Third, recent work[44] shows that Muslims underperform (in education and health) in areas with precolonial Islamic states (for example, in Northern Nigeria and Cameroon and Senegal) owing to weak penetration of the colonial state and limited public-goods investments by missionaries. By contrast, in areas with modest Muslim communities, religious competition pushed Muslims (elites) to adopt Western education. Figure 4d looks at the association between the IM gaps and religious composition. Diversity, as captured by the Herfindahl index, is not a significant correlate. But the religious upward IM gaps are strongly correlated with segregation. The Christian–Muslim upward IM gap is significantly higher (lower) in predominantly Muslim (Christian) regions; Muslims underperform in districts where they are majorities. Evidently, the share of Muslims in the district is the strongest correlate of the Christian–Muslim upward IM differences. Figure 5 further unpacks this association. Figure 5a plots the Christian–Muslim gap in upward IM against the fraction of Muslims across districts, whereas Fig. 5b conditions on the religion-specific literacy rates of the older generation in the district that correlates strongly with IM. The strong positive association remains intact. Figure 5c,d plot the correlation between upward IM and own-religion share separately for Christians and Muslims, conditioning on the literacy of the elders. The likelihood that Christian children of parents without much education will complete primary school is similar in places with small, modest and prominent Christian communities. By contrast, completion of primary education for Muslims is high in regions with small Muslim communities but (very) low in (predominantly) Muslim districts. The negative association between educational opportunity and own-religion share for Muslims echoes US-based evidence that African American children underinvest in education in segregated communities and ghettos. Besides, it squares with India-based results[36], showing a negative association between caste segregation and relative educational mobility.

Fig. 5: District upward IM (gap) and Christian–Muslim population share.



a,**b**, Binned scatterplots of the gap in upward IM between Christians and Muslims (**a**) against the share of Muslims in the district's population conditional on country fixed effects (**b**). Panel **b** also conditions on the religion-specific shares of the older generation with completed primary education in the district. **c**,**d**, Binned within-country scatterplots of upward IM for Christians (**c**) and Muslims (**d**) against their own religious share in the district's population, conditioning on the share of completed primary education of the older generation in the district for Christians and Muslims, respectively. The figure also shows 95% confidence bands obtained by a simple bootstrap procedure. Specifically, for 10,000 bootstrap iterations, we resample the data with replacement, re-estimate the regression in the bootstrap sample and record the estimated coefficient values. We then use the 10,000 bootstrap estimates to predict the dependent variable along a regularly spaced grid from the minimum to the maximum of the independent variable. For each grid point, the lower end of the confidence band is the 2.5th percentile and the upper end is the 97.5th percentile of the 10,000 bootstrap predictions.

Taking stock

We arrived at two takeaways. First, for all faiths, upward IM is higher and downward IM is lower in more developed regions, closer to the capital and the coast, with relatively more colonial investments. However, as Muslims and adherents of Traditional religions reside in less developed, more remote regions with less colonial infrastructure, they are at a disadvantage. Second, religious segregation seems instrumental (although the correlations do not identify causal effects). Christians do well independently of residence, whereas Muslim children underperform compared with other coreligionists in areas where Muslims seem in greater numbers, even when we account for differences in the literacy of the older coreligionists in each district.

Internal migration

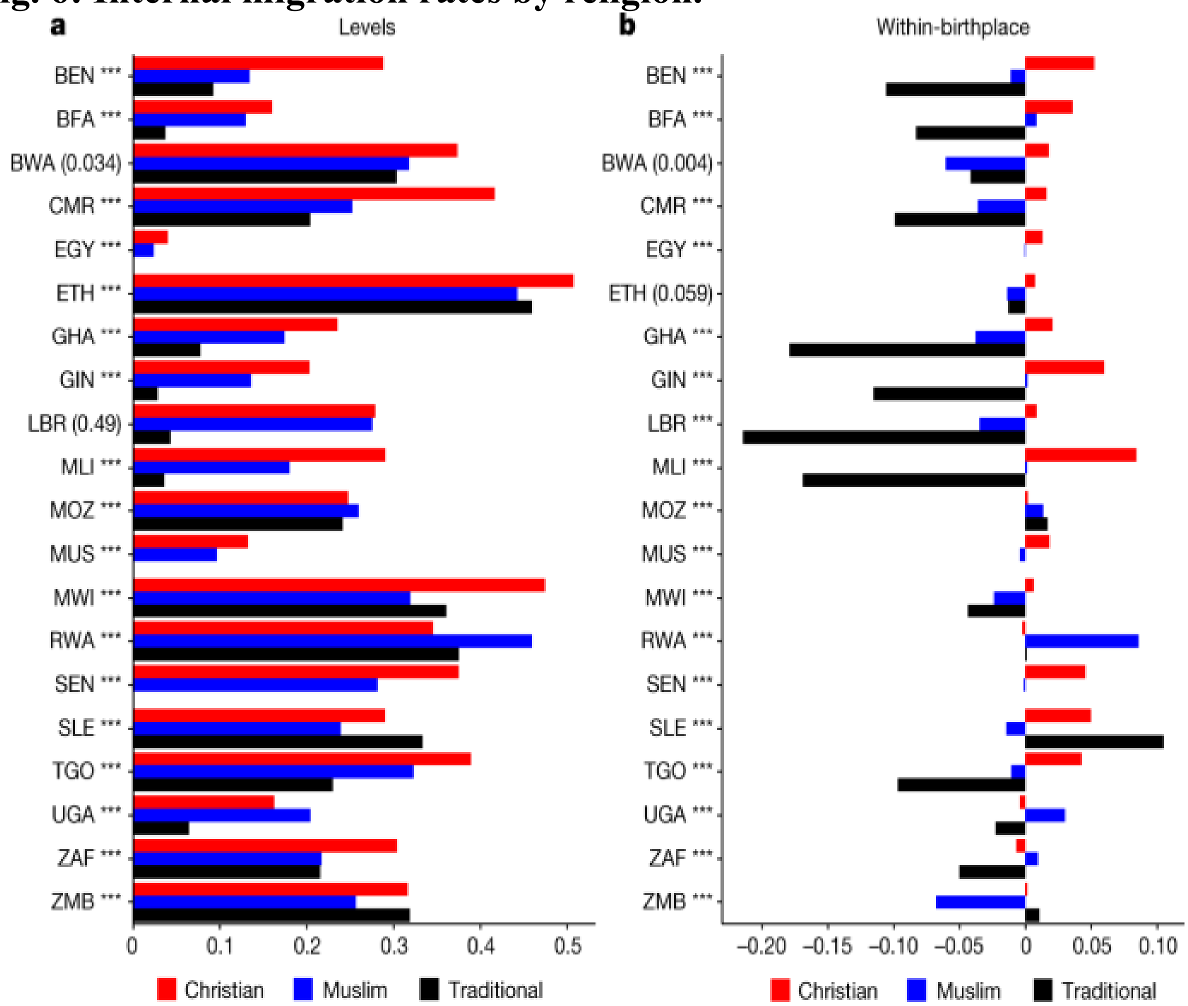
Motivated by the evidence on regional childhood exposure effects for both Muslims and Christians (Fig. 3) and the significant correlation between segregation and religious IM gaps (Fig. 5), we zoom in on internal migration.

Differential internal migration

We tabulate internal migration rates by religious affiliation. We classify as migrants those individuals who, at the time of the census, reside elsewhere than their birthplace district. Figure 6a plots the internal migrant shares for Christians, Muslims and Traditionalists, pooling all censuses across 20 countries (data unavailable for Nigeria). In Fig. 6a and 6b we indicate whether the probability of migration is significantly different between Christians and Muslims in each country. In 17 countries, Christians move at higher rates than Muslims. On average the propensity of Christians to migrate is 0.298 compared to 0.222 for Muslims and 0.194 for Traditionalists. In Cameroon, 40% of Christians reside somewhere other than their birthplace district, whereas the corresponding share for Muslims is 25%. In Ethiopia, the Christian–Muslim difference in emigration is 7 percentage points, whereas in Malawi it is 15 percentage points. The emigration rates of Muslims exceed those of Christians only in Rwanda (by 10.8 percentage points), Uganda (by 4.3 percentage points) and Mozambique

(by 0.9 percentage points). Migration decisions reflect the associated costs and benefits of doing so that may differ across religious lines. Muslims probably face higher migration costs as they reside in relatively remote regions with limited investments. Thus, in Fig. 6b, we report internal migration shares, netting out the mean at the individual's birthplace (weighted by the population of the region in the country) to account for interreligious differences in residence (Extended Data Fig. 3). Differences in internal migration are evident, even when we compare individuals born in the same district, with Christians being 3.6 percentage points more likely to move out than Muslims. Only in Uganda, South Africa, Mozambique and Rwanda, where Muslims are in the minority, are they on average 4 percentage points more likely to move from their birth region compared to Christians. Which other factors shape the uncovered differences, economic, cultural or institutional (interacted with religion), remains uncertain[45,46,47].

Fig. 6: Internal migration rates by religion.

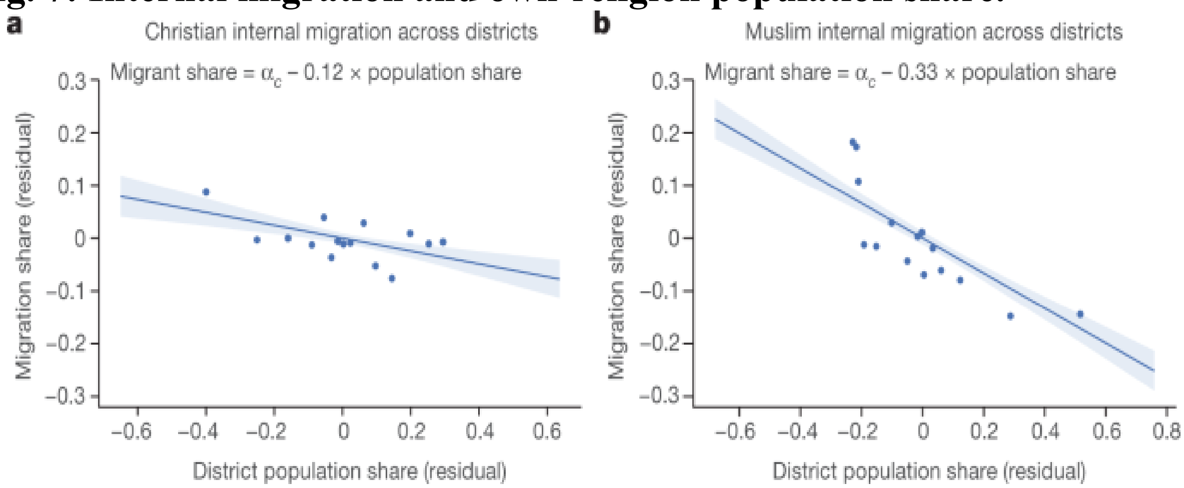


a, Mean values for internal migration for Christians (red bars), Muslims (blue bars) and Africans adhering to Traditional religions (black bars) for each country, pooling all available censuses. A migrant is an individual residing at the time of the census in a district different from their birth region. **b**, Internal migration shares for Christians, Muslims and Traditionalists, netting out the mean value of the individual's birthplace. To get the demeaned-at-the-birth-district statistics, we proceed as follows. First, we get the total weights for individuals in a given country cohort census religion summing across all birthplace districts. Second, we calculate the number of district observations in a given country district cohort census major religion. Third, we divide the district (step 2) with the total (step 1) number of observations to get the district share. Fourth, we multiply the migrant share, individuals residing in other than the birthplace district, to get a weighted demeaned migrant share (step 3). Fifth, we sum the weighted migrant shares across all origin districts for each religious group in each country cohort. Sixth, we take the average across years, as there may be more than one census. The *P* values of the difference in internal migration between Christians and Muslims appear in parentheses; $^{***}P < 0.001$. The *P* values are computed as $2(1-\Phi (|\frac{{\mu }_{{\rm{C}}-{\rm{M}}}^{{\rm{m}}{\rm{i}}{\rm{g}}}}{{\sigma }_{{\rm{C}}-{\rm{M}}}^{{\rm{m}}{\rm{i}}{\rm{g}}}}|)),$ where $\Phi(\,)$ is the standard normal cumulative distribution and ${{\mu }}_{{\rm{C}}-{\rm{M}}}^{{\rm{mig}}}$ and ${\sigma }_{{\rm{C}}-{\rm{M}}}^{{\rm{mig}}}$ are the mean and standard deviation of the distribution of Christian − Muslim migrant share differences. The distribution of differences is computed from 1,000 bootstrap samples for each country, where each sample is a draw with replacement from the original data of the same size as the original data. BEN, Benin; BFA, Burkina Faso; BWA, Botswana; CMR, Cameroon; EGY, Egypt; ETH, Ethiopia; GHA, Ghana; GIN, Guinea; LBR, Liberia; MLI, Mali; MOZ, Mozambique; MUS, Mauritius; MWI, Malawi; RWA, Rwanda; SEN, Senegal; SLE, Sierra Leone; TGO, Togo; UGA, Uganda; ZAF, South Africa; ZMB, Zambia.

Migration and religious segregation

Motivated by the low educational mobility of Muslims in predominantly Muslim districts, we explored the association between migration propensity and religious segregation. Figure 7 illustrates the patterns, plotting internal migration rates for Christians and for Muslims against their own-religion population shares. The (within-country) association is negative for both religious groups but starker for Muslims. The correlation between internal migration and own population share in the birthplace is three times larger for Muslims (coefficient (standard error) −0.33 (0.0248)) than for Christians (estimate (standard error) −0.12 (0.0234)). Muslims, compared to Christians, have a much lower propensity to move out of regions with sizable Muslim communities, exacerbating their initial educational disadvantage.

Fig. 7: Internal migration and own-religion population share.



a,**b**, Binned scatterplots of internal migration rates for Christians (**a**) and Muslims (**b**) against their own religious share in the district's population, conditional on country fixed effects. Also shown are the 95% confidence bands obtained by a simple bootstrap procedure. Specifically, for 10,000 bootstrap iterations, we resample the data with replacement, re-estimate the regression in the bootstrap sample and record the estimated coefficient values. We then use the 10,000 bootstrap estimates to predict the dependent variable along a regularly spaced grid from the minimum to the maximum of the independent variable. For each grid point, the lower end of the confidence band is the 2.5th percentile and the upper end is the 97.5th percentile of the 10,000 bootstrap predictions.

Discussion

We construct religion-specific educational mobility measures since independence across African countries and regions and explore their origins. Three regularities emerge. First, there are significant differences in IM between Christians and Muslims, even comparing Africans living in the same district, with similarly (un)educated elders in their religious group, in households with comparable size and structure and with household heads in the same broad sector and occupation. Second, although Muslims benefit as much as Christians when they move early in life to high-mobility regions, they seem less likely to do so. The comparatively low internal mobility of Muslims accentuates their educational IM deficit, as they (and Traditionalists) reside in less urbanized areas, far from the capital and the coastline areas with limited infrastructure. Third, in areas with large Muslim communities, the Christian–Muslim IM gap is greatest; these highly segregated areas also have the lowest emigration rates among Muslims.

Our study begets more questions than it answers. First, as there is a great deal of variation within faiths, research should explore within-denomination variation distinguishing, for example, between Maliki and Shafi Suni, Ahmadis and Shia Muslims and between Protestants, Copts and Catholics. Doing so would allow delving into the probable causes, teasing apart the role of (1) social norms, (2) faith-specific schooling infrastructure, including *maktabs* and *madrasas* for Muslims and (3) religious leaders and their interaction with state institutions[24,44,45]. Second, our measures of intergenerational mobility in education do not capture how much learning takes place in schools and recent studies stress the low quality of schooling in the continent[48,49]. This may partially rationalize why the first-order differences in education between African Christians and Muslims do not translate into a stark interfaith gap in well-being and occupational specialization. It also highlights the need to estimate faith-specific private and social returns to schooling, both actual and perceived[50], in religiously segregated labour markets with denomination-specific risk-sharing institutions. The voluminous literature we review in Supplementary Information, section A1, documenting higher returns to primary education in low-income settings and Africa in particular[5,6], has paid little attention to the role of religion. Third, as millions are moving to Africa's new megacities, research should explore, and policy-makers should rethink, potential heterogeneity along religious lines of the economic return to migration,

linking it with migration costs and labour markets both at the origin and the destination. Finally, as international institutions and African governments invest heavily in education with school construction programmes, abolishing fees and expanding access and georeferenced data on schools, educational quality and learning become available[40,51], future work on educational reforms needs to carefully explore the roots of inequalities in the take-up of educational policies by religious groups[25].

Methods

Data and educational IM measures

Data reporting

No experiments were performed. No statistical methods were used to predetermine the sample size of the harmonized census data provided by IPUMS. Supplementary Information, section B1 details our procedure for going from the IPUMS data to the sample used in the empirical analysis.

Data

IPUMS International records religious affiliation in 45 censuses from 20 African countries plus Nigeria, for which data come from household surveys between 2006 and 2010. The spatial disaggregation allows us to analyse a maximum of 2,304 districts, typically admin-2 or admin-3 regions. The number of districts used to construct faith-specific IM measures varies by cohort and whether one imposes restrictions in terms of sampling. (See also Supplementary Tables C4 and C5 and the discussion in Supplementary Information, section C3) Extended Data Table 1 reports religious shares by country. Egypt, Senegal, Mali and Guinea are predominantly Muslim (shares exceeding 85%), as well as Sierra Leone (77%) and Burkina Faso (59%). The share of those following Traditional religions is substantial in West Africa, Togo (29%), Benin (22%) and Burkina Faso (19%). When we weigh by the population of a country to account for the limited observations from Nigeria, Christians are about 50%, Muslims 42.7% and Traditionalists about 3% (Extended Data Fig. 1). We discuss conversion dynamics and their

implications for IM in educational attainment estimates in Supplementary Information, section B3.

Sample representativeness

In 2020, the 21 countries hosted roughly 750 million of Africa's 1.35 billion people. North Africa is under-represented, as we have data only from Egypt. The sample includes both relatively rich, educated, with strongly institutionalized countries (for example, South Africa and Botswana with gross domestic product (GDP) per capita of about US $4,000–4,500 in 1995) and relatively poor, weakly institutionalized countries (for example, Ethiopia, Malawi and Mozambique with GDP per capita of about US $250 in 1995). The sample includes former British (Nigeria, Sierra Leone and Malawi), French (Burkina Faso, Senegal and Guinea), German-Belgian (Rwanda) and Portuguese (for example, Mozambique) colonies and protectorates, besides Liberia and Ethiopia. The sample also includes low-state capacity countries with lasting civil wars (Sierra Leone, Mozambique, Rwanda, Liberia and Ethiopia) and more stable ones (South Africa and Botswana). Supplementary Information, section B5 shows that the 21 countries are representative of the continent.

Intergenerational mobility

Our measures of educational IM reflect how 14–18-year-old children fare vis-a-vis cohabiting older generation members, typically biological parents, using primary school completion as the critical educational milestone (Supplementary Fig. A2).

Absolute upward IM in education: $IM_up_{igbcrt} = 1$ in the case of child *i*, of religious affiliation *g*, born in decade *b* in country *c* (and residing in region *r*), observed in census-year *t*, born to parents who have not completed primary schooling completes primary education.

Absolute downward IM in education: $IM_down_{igbcrt} = 1$ in the case of child *i* whose parents have completed primary schooling does not complete primary education.

The IPUMS codebooks suggest that attendance at Christian and Islamic schools is accounted for as long as they do not solely cover religious topics.

Educational dynamics and IM

The literature on IM uses various statistics[52], like (one minus) the intergenerational coefficient obtained from a regression of children on parental schooling[53], rank–rank coefficients and rank movements[29]. Other studies[54,55,56,57] focus, as we do, on absolute transitions. Absolute mobility reflects both overall increases in education over time and movements in the distribution; hence relative and absolute IM measures are not necessarily correlated. See refs. [58,59] for a discussion on the link between absolute IM, relative IM, inequality and growth. We compile new statistics and study absolute upward and downward IM, using primary school completion as the educational cutoff. Focusing on the differences in upward and downward IM across denominations sheds light on the steady-state (ss) differences in educational achievement across religious groups. We can express the evolution of the share of completed primary education for birth cohort $b + 1$, of a religious group g, $\phi_{g,b+1}$, as a function of the share of those with completed primary education in the previous birth cohort, $\phi_{g,b}$, and the rates of upward IM, $u_{g,b}$, and downward IM, $d_{g,b}$.

$$\begin{array}{c}{\varphi }_{g,b+1}={\varphi }_{g,b}(1-{d}_{g,b})+(1-{\varphi }_{g,b}){u}_{g,b}\;\iff \;\varDelta {\varphi }_{g,b+1}\\ \,\,=\;{u}_{g,b}-{\varphi }_{g,b}({u}_{g,b}+{d}_{g,b})\iff {\varphi }_{g,{\rm{ss}}}=\frac{{u}_{g,b}}{{u}_{g,b}+{d}_{g,b}}.\end{array}$$

Cohabitation

We need to observe children's and parental education to build IM statistics. To maximize coverage, we use the average attainment of individuals one generation older than the child in the household. (The results are similar when we take the minimum or maximum or only the father's or mother's education). Matching young individuals to cohabiting older generation members raises concerns, as the transmission of education may differ for children living with and without older family member(s) and across

religions. We focus on individuals aged 14–18 years, as primary education is mostly complete by then and cohabitation rates are high[57]. Supplementary Information, section B2 gives details. Cohabitation rates with older generation relatives, mainly biological parents and sometimes uncles and aunts, hover between 82% and 91%, without much difference between Christians, Muslims and Traditionalists.

Religious affiliation across generations

We explore the transmission of religious affiliation from parents to 14–18 year-old children. On the one hand, there is high intergenerational inertia for both Muslims and Christians. The likelihood that children of Muslim or Christian parents will report a different creed is, on average, less than 3%; in most countries, it is less than 1%. These estimates are close to the ones reported by ref. [60] across 19 African countries. On the other hand, it is common for African Muslims and Christians to follow traditional religious rituals and ceremonies[15]. However, the Census does not record 'mixed/dual' religious affiliation. To the extent that educated Africans adhering to Traditional religions alongside Christianity or Islam will respond that they are Christians or Muslims, the upward IM estimates for Traditionalists will be underestimated. Supplementary Information, section B3 gives details and graphical illustrations of the conversion dynamics across denominations and discusses their implications for our patterns.

Ethnicity and religion

Given the voluminous research on ethnicity in Africa[16,17,18], we examine the interplay between religion and ethnicity by tabulating censuses in which both are recorded. Religion transcends ethnicity (Supplementary Information, section B4). Although a few ethnicities are monoreligious (for example, Wolof and Fula in Senegal and the Somali in Ethiopia are Muslim and the Agew in Ethiopia and the Acholi in Uganda are Christian), most ethnicities, large and small, are multireligious. For example, the Oromo in Ethiopia, the Yoruba in Nigeria and the Sena in Mozambique are split between Christianity and Islam, whereas many groups in West Africa are split between Christianity and Traditional religions. There are

dozens of ethnicities split between Traditional religions, Islam and Christianity, like the Gurma, the Basari and the Goulmancema in West Africa. Supplementary Figure D3 reports the religious IM gaps leveraging both cross-ethnicity and within-ethnicity variation (across individuals for whom IPUMS records ethnicity). (See also Supplementary Information, section D2.2).

Explaining religious educational IM gaps

Methodology

To arrive at the Christian–Muslim and the Christian–Animist gaps in upward IM and downward IM in Fig. 2, we estimate the following regression with OLS:

$${{\rm{I}}{\rm{M}}}_{ibchdt}^{{\rm{rel}}}={\alpha }_{cb}+{\gamma }_{m}{\rm{Muslim}}+{\gamma }_{a}{\rm{Animist}}+{\delta }_{h}{{\bf{H}}}_{h}^{{\prime} }+{\theta }_{h}{{\bf{I}}}_{h}^{{\prime} }+\psi {{\bf{E}}}_{e,r,b-1}^{{\prime} }+{\varphi }_{d,u/r}+{{\epsilon }}_{ibchdt}.$$

The dependent variable denotes upward or downward IM for child *i* of religious affiliation rel, born in decade *b*, in household *h*, residing in district *d* in country *c*, recorded in census *t*. 'Muslim' is an indicator for adherents to Islam; 'Animist' identifies children of Traditionalists. ${{\bf{H}}}_{h}^{{\prime} }$ is a vector of household features, including size, composition and family organization. ${{\bf{I}}}_{h}^{{\prime} }$ reflects occupation and industry indicators for the older generation in the household. ${{\bf{E}}}_{e\,,r\,,b-1}^{{\prime} }$ denotes the share of the older generation with completed primary education in the district for each religious group. ${{\boldsymbol{\varphi }}}_{d,u/r}$ is a vector of district-specific constants interacted with an urban indicator. Parameters γ_m and γ_a reflect the educational gap of Muslims and Traditionalists vis-a-vis Christians, the omitted category.

Childhood regional exposure effects

Methodology

To isolate regional childhood exposure effects from spatial sorting, we use an approach[37] that exploits differences in the timing of children's moves across districts with different levels of upward IM, adjusting it to derive religion-specific exposure[38]. The regional childhood exposure effects, reported in Fig. 3, are estimated from the following OLS specification:

$$\begin{array}{c}{\text{IM_up}}_{ihbmcod}^{{\rm{rel}}}={\alpha }_{ob}+{\alpha }_{m}+\mathop{\sum }\limits_{m=1}^{18}{{\rm{\beta }}}_{m}^{{\rm{rel}}}\times I({m}_{i}=m)\times \,{\varDelta }_{odb}^{{\rm{nm}}}\\ \,\,\,+\mathop{\sum }\limits_{b={b}_{0}}^{B}{k}_{b}\times I({b}_{i}=b)\times {\varDelta }_{odb}^{{\rm{nm}}}+{{\epsilon }}_{ihbmcod}\end{array}$$

The specification relates primary education completion for child *i*, from household *h*, of birth cohort *b*, whose parents have not completed primary school, who moved from birthplace district *o* in country *c* to destination district *d* at age *m* in the same country, to differences in upward IM between origin and destination, among non-movers of the same cohort $b({\varDelta }_{{odb}}=\overline{{\rm{IM}}}{\text{_up}}_{{bd}}^{nm}-\overline{{\rm{IM}}}{\text{_up}}_{{bo}}^{nm}\,)$. $\overline{{\rm{IM}}}_{{\rm{up}}}_{b}^{{\rm{nm}}}$ summarizes the economic, social and institutional environment which shapes educational mobility in a district. We construct an overall measure of origin–destination differences in upward IM (${\varDelta }_{odb}^{{\rm{all}}}$) and a religion-specific one (${\varDelta }_{odb}^{{\rm{rel}}}$). Origin-region × birth-decade fixed effects, α_{ob}, account for unobserved factors of the child's cohort and birthplace. The specification also includes interactions of destination–origin cohort IM differences with cohort effects to account for potential differential measurement error across cohorts (this has no effect). The parameters of interest, ${\beta }_{m}^{{\rm{rel}}}$, capture how children's attainment varies with the age of their move to districts with higher or lower upward IM, conditional on age-of-move constants, α_m, which absorb disruption and other age-specific features affecting education. If regions matter for mobility, the earlier the move, the greater the impact. As we include origin cohort specific constants, we leverage variation among

children born in the same district and decade, moving to regions with different educational mobility. Differences in the age-of-move slopes, ${\gamma }_{m}={\beta }_{m}^{{\rm{rel}}}-{\beta }_{m+1}^{{\rm{rel}}}$, reflect the impact of an extra year in the high-mobility district, regional childhood exposure effects.

Sample

For the implementation of the movers' design in the section on Sorting and childhood regional exposure (Fig. 3) that teases apart childhood exposure regional effects from spatial sorting, we need data not only on the district of birth and residence but also on the length of stay in the current location. IPUMS provides such information for 13 countries (Supplementary Table B1). The 13 countries come from all African regions. (1) Southern Africa: we have data from three (out of five) countries in the southern part of the continent, South Africa, Zambia and Malawi, missing Mozambique and Botswana. (2) Western Africa: the childhood exposure effects are estimated using data from six (out of ten) West African countries, Benin, Ghana, Guinea, Togo, Mali and Cameroon. (3) East Africa: we have all countries from Eastern Africa (Uganda, Rwanda and Ethiopia) except Mauritius. (4) North Africa: the mover's design includes Egypt, the only North African country. The 13-country sample includes French and British colonies and covers relatively poor (with meagre education) nations and more advanced economies.

Regional correlates of interreligious educational mobility differences

Religious differences in residence

We explore differences in residence attributes among adherents of the three main religions, running country–birth cohort fixed-effects regressions associating geographic/location, at-independence development and historical features to indicator variables for Muslims and Traditionalists, respectively, with Christians serving as the omitted category. The specification reads:

$$\begin{array}{l}{Y}_{i,c,b,r}={\alpha }_{c,b}+{\psi }_{1}{{\rm{Muslim}}}_{i,c,b,r}+{\psi }_{2}{{\rm{Traditional}}}_{i,c,b,r}+{\zeta }_{i,r,c,b}.\end{array}$$

Extended Data Figure 3 reports population-weighted least-squares estimates that reflect average differences in the respective outcomes, $Y_{i,c,b,r}$, between Christians and Muslims and Christian and Traditionalists partialling out country cohort constants, $\alpha_{c,b}$. Our exploration relates to research on the spread of Christianity and Islam in Africa. The following patterns emerge when we compare residence attributes between Muslims and Christians: (1) Muslims reside in less developed and densely populated regions, are more reliant on agriculture and are somewhat less urbanized; (2) Muslims reside in regions away from the capitals and the coastline, in line with works showing that missionaries mostly settled along the coast; and (3) Muslims reside in districts further away from colonial roads and railroads and far from Protestant and Catholic missions. Turning to Animists, the tabulations show the following: (1) similar to Muslims, Africans adhering to Traditional religions reside in less densely, more rural, agriculture-oriented regions; (2) adherents of Traditional religions reside in districts even further away from the capitals than Muslims; (3) Animists are more likely to settle in malaria-prevalent districts; and (4) Traditionalists are found in districts far from the colonial infrastructure.

Specification of regional correlates of religious IM gap

The specification on the correlates of regional differences in upward IM between Christians and Muslims and Christians and Traditionalists reads:

$$\begin{array}{c}\mathop{{\overline{{\rm{IM}}}}_{r,c,b}^{{\rm{C}}}}\limits^{}-\mathop{{\overline{{\rm{IM}}}}_{r,c,b}^{{\rm{M,T}}}}\limits^{}={\gamma }_{{\rm{c}}}+{\varphi }_{1}{D}_{r,c}+{\varphi }_{2}{G}_{r,c}+{\varphi }_{3}{H}_{r,c}+{\varphi }_{4}{R}_{r,c}+\lambda {E}_{r,c}^{{\rm{old}}}+{{\epsilon }}_{r,c,b}.\end{array}$$

The dependent variable is the difference in upward IM between Christians and Muslims and between Christians and Traditionalists born in decade

(birth cohort) b, in region r, in country c. (Hats denote regional averages across birth cohort regions). The explanatory variables are regional proxies of early (at independence) development (D), geography-location (G), historical aspects (H) and religious composition (R), which we include one by one in the empirical model as our objective is to characterize regional religious IM differences (rather than identify causal effects). As the specifications include country constants (γ_c), the coefficients capture the within-country correlation.

Specification of regional correlates of IM, by religion

We also estimate the regional specifications separately for Muslims, Christians and adherents of Traditional religions and compared the coefficient estimates of the location, early development and historical and compositional statistics. Extended Data Figure 4 reports the regional correlates of mobility separately for adherents of each religious group. Extended Data Figure 5 reports otherwise similar specifications, also controlling for the share of the older generation with completed primary education in the district that is the strongest correlate of educational mobility.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All newly built statistics on faith-specific absolute upward and downward educational IM by country, region and sex are based on census data compiled, processed and harmonized by IPUMS. IPUMS International microdata are publicly available free of charge. To access them, the prospective user may submit an electronic authorization form providing name, electronic address and institutional affiliation here: https://international.ipums.org/international/. Because our analyses are based

on secondary, de-identified, publicly available data, we do not have an IRB waiver.

Code availability

The code used to construct and analyse the data was written in R v.4.2.2 and Python v.3.11. The replication code and data files are available on https://github.com/imreligionafrica/imreligionafrica

References

1. Nkrumah, K. *Consciencism* (Monthly Review Press, 1964).

2. Nyerere, J. K. *Freedom and Development* (Oxford Univ. Press, 1973).

3. *World Development Report 2018: Learning to Realize Education's Promise* (World Bank, 2018).

4. Wodon, Q., Male, C., Montenegro, C., Nguyen, H. & Onagoruwa, A. *Educating Girls and Ending Child Marriage: A Priority for Africa* (World Bank, 2018).

5. Psacharopoulos, G. & Patrinos, H. Returns to investment in education: a decennial review of the global literature. *Edu. Econ.* **26**, 445–458 (2018).

6. Montenegro, C. E. & Patrinos, H. A. *Comparable Estimates of Returns to Schooling Around the World* (World Bank, 2014).

7. *Religion and Education Around the World* (Pew Research Center, 2016).

8. Nathan, N. Religious conversion in colonial Africa. *Am. Econ. Rev.* **100**, 147–152 (2010).

9. Jedwab, R., Meier Zu Selhausen, F. & Moradi, A. *The Economics of Missionary Expansion: Evidence from Africa and Implications for*

Development (Univ. of Oxford, 2018).

10. Julia, C. & Valeria, R. The long-term effects of the printing press in sub-Saharan Africa. *Am. Econ. J. Appl. Econ.* **8**, 69–99 (2016).

11. Majgaard, K. & Mingat, A. *Education in Sub-Saharan Countries: A Comparative Analysis* (World Bank, 2011).

12. Abernerthy, D. B. *The Political Dilemma of Popular Education: An African Case* (Stanford Univ. Press, 1969).

13. Fafunwa, B. A. *History of Education in Nigeria* (Routledge, 1964).

14. Laitin, D. D. *Hegemony and Culture: Politics and Religious Change among the Yoruba* (Univ. of Chicago Press, 1986).

15. Peel, J. D. Y. *Aladura*: *A Religious Movement Among the Yoruba* (Oxford Univ. Press, 1968).

16. Posner, D. N. *Institutions and Ethnic Politics in Africa* (Cambridge Univ. Press, 2005).

17. Bates, R. H. *Markets and States in Tropical Africa: The Political Basis of Agricultural Policies* (Univ. of California Press, 2005).

18. Christensen, D. & Laitin, D. D. *African States Since Independence: Order, Development, and Democracy* (Yale Univ. Press, 2019).

19. Guiso, L., Sapienza, P. & Zingales, L. People's opium? Religion and economic attitudes. *J. Monet. Econ.* **50**, 228–282 (2003).

20. Becker, S. O., Rubin, J. & Woessmann, L. in *The Handbook of Historical Economics* (eds Bisin, A. & Federico, G.) 585–639 (Elsevier, 2021).

21. Kuran, T. Islam and economic performance: historical and contemporary links. *J. Econ. Lit.* **56**, 1292–1359 (2018).

22. Manglos-Weber, N. D. Identity, inequality, and legitimacy: religious differences in primary school completion in sub-Saharan Africa. *J. Sci. Study Relig.* **56**, 302–322 (2017).

23. Platas, M. R. *The Religious Roots of Inequality in Africa*. PhD thesis, Stanford Univ. (2016).

24. Platas, M. R. *Culture and the Persistence of Educational Inequality: Lessons from the Muslim–Christian Education Gap in Africa* (New York Univ., 2018).

25. Evans, D. K. & Acosta, A. M. Education in Africa: what are we learning? *J. African Econ.* **30**, 13–54 (2021).

26. Narayan, A. et al. *Fair Progress?: Economic Mobility Across Generations Around the World* (World Bank, 2018).

27. Hertz, T. in *Unequal Chances: Family Background and Economic Success* (eds Bowles, S. et al.) 165–191 (Princeton Univ. Press, 2005).

28. Chetty, R., Hendren, N., Kline, P. & Saez, E. Where Is the land of opportunity? The geography of intergenerational mobility in the United States. *Q. J. Econ.* **129**, 1553–1623 (2014).

29. Geng, Y. *Intergenerational Mobility in China across Space and Time* (Univ. Pompeu Fabra, 2018).

30. Alesina, A., Hohmann, S., Michalopoulos, S. & Papaioannou, E. Intergenerational mobility in Africa. *Econometrica* https://doi.org/10.3982/ECTA17018 (2021).

31. Wantchekon, L., Klasnja, M. & Novta, N. Education and human capital externalities: evidence from colonial Benin. *Q. J. Econ.* **130**, 703–757 (2015).

32. Carillo, B., Charris, C. & Iglesias, W. Moved to poverty? A legacy of apartheid in South Africa. *Am. Econ. J. Appl. Econ.* (in the press).

33. Asher, S., Novosad, P. & Rafkin, C. *Intergenerational Mobility in India: Estimates From New Methods and Administrative Data* (World Bank, 2018).

34. Michalopoulos, S., Naghavi, A. & Prarolo, G. Trade and geography in the spread of Islam. *Econ. J.* **128**, 3210–3241 (2018).

35. Alesina, A. & Zhuravskaya, E. Segregation and the quality of government in a cross section of countries. *Am. Econ. Rev.* **101**, 1872–1911 (2011).

36. Asher, S., Novosad, P. & Rafkin, C. *Intergenerational Mobility in India: New Methods and Estimates Across Time, Space, and Communities* (2020); http://paulnovosad.com/pdf/anr-india-mobility.pdf.

37. Chetty, R. & Hendren, N. The impacts of neighborhoods on intergenerational mobility I: childhood exposure effects. *Q. J. Econ.* **133**, 1107–1162 (2018).

38. Chetty, R., Hendren, N., Jones, M. R. & Porter, S. Race and economic opportunity in the United States: an intergenerational perspective. *Q. J. Econ.* **135**, 711–783 (2020).

39. Finkelstein, A., Gentzkow, M. & Williams, H. Place-based drivers of mortality: evidence from migration. *Am. Econ. Rev.* **111**, 2697–2735 (2021).

40. Walter, T. F. *Misallocation of State Capacity? Evidence from Two Million Primary Schools*. PhD thesis, London School of Economics and Political Science (2020).

41. Michalopoulos, S. & Papaioannou, E. Historical legacies and African development. *J. Econ. Lit.* **58**, 53–128 (2020).

42. Cutler, D. M. & Glaeser, E. L. Are ghettos good or bad? *Q. J. Econ.* **112**, 827–872 (1997).

43. Cutler, D. M., Glaeser, E. L. & Vigdor, J. L. When are ghettos bad? Lessons from immigrant segregation in the United States. *J. Urban Econ.* **63**, 759–774 (2008).

44. Bauer, V., Platas, M. R. & Weinstein, J. The historical legacies of Islamic rule in Africa: colonial responses and contemporary development. *World Dev.* **152**, 105750 (2022).

45. Acemoglu, D. & Robinson, J. A. *Culture, Institutions and Social Equilibria: A Framework* (NBER, 2021); https://www.nber.org/papers/w28832

46. Bisin, A. & Verdier, T. *On the Joint Evolution of Culture and Institutions* (NBER, 2017); https://www.nber.org/papers/w23375

47. Munshi, K. & Rosenzweig, M. Networks and misallocation: insurance, migration, and the rural–urban wage gap. *Am. Econ. Rev.* **106**, 46–98 (2016).

48. Singh, A. Learning more with every year: school year productivity and international learning divergence. *J. Euro. Econ. Assoc.* **18**, 1770–1813 (2020).

49. Sandefur, J. Internationally comparable mathematics scores for fourteen African countries. *Econ. Educ. Rev.* **62**, 267–286 (2018).

50. Jensen, R. The (perceived) returns to education and the demand for schooling. *Q. J. Econ.* **125**, 515–548 (2010).

51. Angrist, N., Djankov, S., Goldberg, P. K. & Patrinos, H. A. Measuring human capital using global learning data. *Nature* **592**, 403–408 (2021).

52. Black, S. E. & Devereux, P. J. in *The Handbook of Labor Economics* Vol. 4B (eds Card, D. & Orley Ashenfelter, O.) 1487–1541 (Elsevier, 2011).

53. Hertz, T. et al. The inheritance of educational inequality: international comparisons and fifty-year trends. *B.E. J. Econ. Anal. Poli.* https://doi.org/10.2202/1935-1682.1775 (2008).

54. Davis, J. & Mazumder, B. *The Decline in Intergenerational Mobility After 1980* (Federal Reserve Bank of Chicago, 2020).

55. Derenoncourt, E. Can you move to opportunity? Evidence from the great migration. *Am. Econ. Rev.* **112**, 369–408 (2022).

56. Chetty, R. et al. Fading American dream: trends in absolute income mobility since 1940. *Science* **356**, 398–406 (2017).

57. Card, D., Domnisoru. C. & Taylor, L. The intergenerational transmission of human capital: evidence from the Golden Age of upward mobility. *J. Labor Econ.* **40**, S39–S95 (2022).

58. Berman, Y. The long run evolution of absolute intergenerational mobility. *Am. Econ. J. Appl. Econ.***14**, 61–83 (2022).

59. Deutscher, N. & Mazumder, B. Measuring intergenerational income mobility: a synthesis of approaches. *J. Econ. Lit. Appl. Econ.* (in the press).

60. *Tolerance and Tension: Islam and Christianity in Sub-Saharan Africa* (Pew Research Center, 2010).

Acknowledgements

The paper is dedicated to A.A., who passed away while we were revising the manuscript. We will never forget his boundless curiosity, drive and charm. We terribly miss his wit, smile and can-do attitude. We thank M. Chatzigakis for his superb assistance with the data and codes and R. Chetty, E. Glaeser, N. Hendren and M. Poschke for their suggestions and comments. We have received helpful feedback and comments from seminar participants at the RIDGE Conference, the NBER Summer Institute, the AEA Conference (2021), the CEPR Conference on The Economics of Religion, King's College, Harvard, Zurich, the New Economic School, Trinity College Dublin, the World Bank, Hebrew University and Pantheon-Sorbonne. We also thank R. Jedwab, A. Storeygard, J. Cage, V. Rueda and N. Nunn for sharing their data. E.P. acknowledges support from the European Research Council (ERC consolidator grant ORDINARY) and the Wheeler Institute for

Business and Development at the London Business School. All errors and omissions are our responsibility.

Author information

Authors and Affiliations

1. Department of Economics, Harvard University, Cambridge, MA, USA

   Alberto Alesina

2. Sihlquai 10, Adliswil, Switzerland

   Sebastian Hohmann

3. Department of Economics, Brown University, Providence, RI, USA

   Stelios Michalopoulos

4. London Business School, London, UK

   Elias Papaioannou

Contributions

A.A., S.H., S.M. and E.P. were the joint principal investigators for this project. A.A. designed the study, supervised the analysis and wrote the first draft of the paper. S.H. designed the study, supervised all parts of the analysis, analysed the data and wrote and edited the paper. S.M. designed the study, supervised all parts of the analysis, analysed the data and wrote and edited the paper. E.P. designed the study, supervised all parts of the analysis, analysed the data, wrote and edited the paper. The study has no local authors as the analysis is based on secondary data, harmonized and made publicly available by IPUMS International.

Corresponding authors

Correspondence to Stelios Michalopoulos or Elias Papaioannou.

Ethics declarations

Competing interests

The authors declare no competing interests.

Peer review

Peer review information

Nature thanks Sudhanshu Handa, David Laitin and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Major Religions. 21 African Countries.

The figure plots the population share of the main religions in our sample of 21 African countries. Panel (a) reports the shares across 82,037,564 individuals of all ages. Panel (b) reports weighted shares using the countries' populations in 1980. The Christian share combines various denominations, like Orthodox, Catholic, and Protestant, available in some censuses. Likewise, the Muslim share combines various branches, like Sunni and Shia. Traditional also combines various indigenous religions, like Vodun, Animist, and Traditional religions. Supplementary Fig. A2 gives country-specific details.

Extended Data Fig. 2 Christian–Muslim IM Gap Drivers, by Country.

The figure portrays how the Christian–Muslim gap in educational upward IM (panel (a)) and downward IM (panel (b)) changes as we add controls for the household structure (model (2), in green), parental occupational specialization, the industry of employment, and rural–urban residence (model (3), in red), the share completed primary education of the older generation of the same religious group in the district (model (4), in purple), and district x rural/urban fixed effects (model (5), in dark yellow) for each country. The last permutation (model (6), in light blue) restricts estimation in half of each country's districts, where differences in completed primary education of the older generation between Christians and Muslims are the smallest. The bars on the top (model (1), in dark blue) reflect the baseline inter-religious differences in IM, conditioning on birth cohort fixed effects and age constants. The figure gives linear probability model (OLS) estimates.

Extended Data Fig. 3 District Correlates of Residence by Religious Affiliation.

The figure plots OLS regression coefficients associating the variable listed on the left of the graph to indicator variables for Muslims and Traditionalists, with Christians serving as the omitted category, conditioning on country–birth cohort fixed effects. The estimates have, therefore, a within-country–birth-decade test of means interpretation. There are three categories of independent variables. (i) Regional proxies of development before independence. (ii) Regional geographic and location features. (iii) Historical variables of colonial investments and precolonial statehood. Supplementary Section F gives variable definitions and sources. Two-standard-error bands based on heteroskedasticity adjusted clustered at the country level are also reported. The point estimates (green and blue dots in the figure) were obtained by running separate regressions of the district-level Christian–Muslim and Christian - Traditional IM gap respectively on each district-level variable (indicated on the vertical axis of the figure). The IM gap is defined as the average IM of Christians minus the average IM of Muslims or Traditionalists in the district. Before running each regression, we

standardize the dependent and independent variable by subtracting its sample mean and dividing by its sample standard deviation.

Extended Data Fig. 4 Regional IM Correlates by Religion, Country Fixed Effects.

The figure plots correlations (standardized "beta" coefficients) between intergenerational mobility (IM) and various regional characteristics for Christians (red star), Muslims (green rhombus), and Traditionalists (blue square). Panel (a) examines upward IM that reflects the likelihood that young individuals, aged 14–18, residing in households where the older generation has not completed primary schooling, will complete primary education. Panel (b) examines downward IM that reflects the likelihood that young individuals, aged 14–18, residing in households where the older generation has completed primary schooling will fail to do so. There are four categories of IM correlates. (i) Proxies of development before independence. (ii) Location and geographic features. (iii) Historical variables, including colonial-era investments and precolonial statehood. (iv) Homogeneity, captured by the shares of each of the three religious groups. Supplementary Section F gives variable definitions and sources. All specifications include country fixed effects (constants not reported). Standard errors are clustered at the country level. The point estimates (green and blue dots in the figure) were obtained by running separate regressions of the district-level Christian–Muslim and Christian - Traditional IM gap respectively on each district-level variable (indicated on the vertical axis of the figure). The IM gap is defined as the average IM of Christians minus the average IM of Muslims or Traditionalists in the district. Before running each regression, we standardize the dependent and independent variable by subtracting its sample mean and dividing by its sample standard deviation.

Extended Data Fig. 5 Regional IM Correlates, Country FE, Cond. on Own Religion Old's Completed Primary.

The figure plots correlations (standardized "beta" coefficients) between intergenerational mobility (IM) and various regional characteristics for Christians (red star), Muslims (green rhombus), and Traditionalists (blue square), conditioning on their own religious group older generation's

completed primary education in the district. Panel (a) examines upward IM that reflects the likelihood that young individuals, aged 14–18, residing in households where the older generation has not completed primary schooling, will complete primary education. Panel (b) examines downward IM that reflects the likelihood that young individuals, aged 14–18, residing in households where the older generation has completed primary schooling will fail to do so. There are four categories of IM correlates. (i) Proxies of development before independence. (ii) Location and geographic features. (iii) Historical variables, including colonial-era investments and precolonial statehood. (iv) Homogeneity, captured by the shares of each of the three religious groups. Appendix Section F gives variable definitions and sources. All specifications include country fixed effects (constants not reported). Standard errors are clustered at the country level. The point estimates (green and blue dots in the figure) were obtained by running separate regressions of the district-level Christian–Muslim and Christian–Traditionalist IM gap, respectively, on each district-level variable (indicated on the vertical axis of the figure). The IM gap is defined as the average IM of Christians minus the average IM of Muslims or Traditionalists in the district. Before running each regression, we standardize the dependent and independent variable by subtracting its sample mean and dividing by its sample standard deviation.

Extended Data Fig. 6 District Correlates of Christian–Muslim and Christian–Traditionalist IM Gaps, Conditional on Older Generation's Completed Primary.

The figure plots correlations (standardized "beta" coefficients) between upward IM gaps of Christians and Muslims (green circle) and Christians and Traditionalists (blue square), averaged across individuals in a district, and various regional characteristics. (i) Proxies of development before independence. (ii) Location and geographic features. (iii) Historical variables of colonial investments and precolonial statehood. (iv) Homogeneity, captured by the shares of each of the three religious groups. Supplementary Section F gives variable definitions and sources. All specifications include country fixed effects (constants not reported). Two-standard-error bands based on heteroskedasticity adjusted clustered at the country level are also reported. The point estimates (green and blue dots in the figure) were obtained by running separate regressions of the district-level

Christian–Muslim and Christian–Traditionalist IM gap respectively on each district-level variable (indicated on the vertical axis of the figure). The IM gap is defined as the average IM of Christians minus the average IM of Muslims or Traditionalists in the district. Before running each regression, we standardize the dependent and independent variable by subtracting its sample mean and dividing by its sample standard deviation.

Extended Data Table 1 Major religion shares by country
Extended Data Table 2 Summary Statistics. District-level Christian–Muslim IM Gaps

Supplementary information

Supplementary Information Supplementary Sections A–F.

Reporting Summary

Peer Review File

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Alesina, A., Hohmann, S., Michalopoulos, S. *et al.* Religion and educational mobility in Africa. *Nature* **618**, 134–143 (2023). https://doi.org/10.1038/s41586-023-06051-2

- Received: 10 September 2021
- Accepted: 04 April 2023
- Published: 17 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06051-2

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Changes in education levels across generations in Africa are linked to religion**
  - Melina R. Platas

Nature (2023)

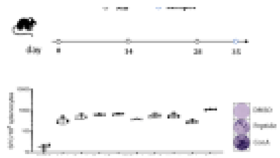
Changes in education levels across generations in Africa are linked to religion

- Melina R. Platas

Nature News & Views 17 May 2023

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06051-2

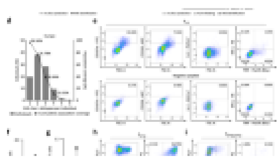
| Section menu | Main menu |



Neoantigen cancer vaccine augments anti-CTLA-4 efficacy

02 February 2022

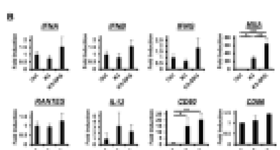
Erika Salvatori, Lucia Lione, … Fabio Palombo



The oncogenic fusion protein DNAJB1-PRKACA can be specifically targeted by peptide-based immunotherapy in fibrolamellar hepatocellular carcinoma

27 October 2022

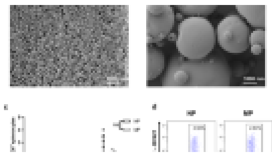
Jens Bauer, Natalie Köhler, … Juliane S. Walz



In situ vaccination using unique TLR9 ligand K3-SPG induces long-lasting systemic immune response and synergizes with systemic and local immunotherapy

08 February 2022

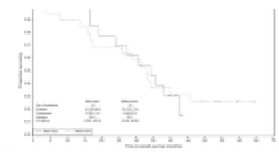
Hirokazu Okada, Ken Takahashi, … Hiroshi Seno



[PLGA-particle vaccine carrying TLR3/RIG-I ligand Riboxxim synergizes with immune checkpoint blockade for effective anti-cancer immunotherapy]

18 May 2021

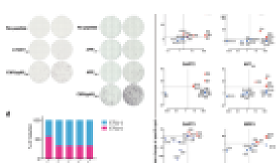
Julia Koerner, Dennis Horvath, … Marcus Groettrup



[TG01/GM-CSF and adjuvant gemcitabine in patients with resected RAS-mutant adenocarcinoma of the pancreas (CT TG01-01): a single-arm, phase 1/2 trial]

17 February 2020

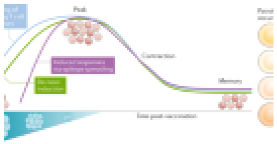
Daniel H. Palmer, Juan W. Valle, … Svein Dueland



[Peptide vaccine-treated, long-term surviving cancer patients harbor self-renewing tumor-specific CD8+ T cells]

03 June 2022

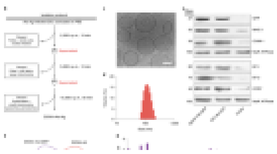
Eishiro Mizukoshi, Hidetoshi Nakagawa, … Shuichi Kaneko



[Advances in the development of personalized neoantigen-based therapeutic cancer vaccines]

20 January 2021

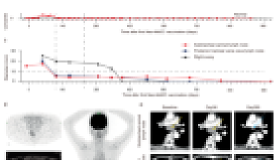
Eryn Blass & Patrick A. Ott



[A nanovaccine for antigen self-presentation and immunosuppression reversal as a personalized cancer immunotherapy strategy]

11 April 2022

Chao Liu, Xue Liu, … Gang Liu



[Durable complete response to neoantigen-loaded dendritic-cell vaccine following anti-PD-1 therapy in metastatic gastric cancer]

03 June 2022

Zengqing Guo, Yuan Yuan, … Xi Zhang

- Article
- [Open Access]

- Published: 10 May 2023

Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer

- Luis A. Rojas,
- Zachary Sethna,
- Kevin C. Soares,
- Cristina Olcese,
- Nan Pang,
- Erin Patterson,
- Jayon Lihm,
- Nicholas Ceglia,
- Pablo Guasp,
- Alexander Chu,
- Rebecca Yu,
- Adrienne Kaya Chandra,
- Theresa Waters,
- Jennifer Ruan,
- Masataka Amisaki,
- Abderezak Zebboudj,
- Zagaa Odgerel,
- George Payne,
- Evelyna Derhovanessian,
- Felicitas Müller,
- Ina Rhee,
- Mahesh Yadav,
- Anton Dobrin,
- Michel Sadelain,
- Marta Łuksza,
- Noah Cohen,
- Laura Tang,
- Olca Basturk,

- Mithat Gönen,
- Seth Katz,
- Richard Kinh Do,
- Andrew S. Epstein,
- Parisa Momtaz,
- Wungki Park,
- Ryan Sugarman,
- Anna M. Varghese,
- Elizabeth Won,
- Avni Desai,
- Alice C. Wei,
- Michael I. D’Angelica,
- T. Peter Kingham,
- Ira Mellman,
- Taha Merghoub,
- Jedd D. Wolchok,
- Ugur Sahin,
- Özlem Türeci,
- Benjamin D. Greenbaum ,
- William R. Jarnagin,
- Jeffrey Drebin,
- Eileen M. O’Reilly &
- …
- Vinod P. Balachandran

Show authors

Nature **volume 618**, pages 144–150 (2023)

- 124k Accesses
- 2438 Altmetric
- Metrics details

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is lethal in 88% of patients[1], yet harbours mutation-derived T cell neoantigens that are suitable for vaccines [2,3]. Here in a phase I trial of adjuvant autogene cevumeran, an individualized neoantigen vaccine based on uridine mRNA–lipoplex nanoparticles, we synthesized mRNA neoantigen vaccines in real time from surgically resected PDAC tumours. After surgery, we sequentially administered atezolizumab (an anti-PD-L1 immunotherapy), autogene cevumeran (a maximum of 20 neoantigens per patient) and a modified version of a four-drug chemotherapy regimen (mFOLFIRINOX, comprising folinic acid, fluorouracil, irinotecan and oxaliplatin). The end points included vaccine-induced neoantigen-specific T cells by high-threshold assays, 18-month recurrence-free survival and oncologic feasibility. We treated 16 patients with atezolizumab and autogene cevumeran, then 15 patients with mFOLFIRINOX. Autogene cevumeran was administered within 3 days of benchmarked times, was tolerable and induced de novo high-magnitude neoantigen-specific T cells in 8 out of 16 patients, with half targeting more than one vaccine neoantigen. Using a new mathematical strategy to track T cell clones (CloneTrack) and functional assays, we found that vaccine-expanded T cells comprised up to 10% of all blood T cells, re-expanded with a vaccine booster and included long-lived polyfunctional neoantigen-specific effector $CD8^+$ T cells. At 18-month median follow-up, patients with vaccine-expanded T cells (responders) had a longer median recurrence-free survival (not reached) compared with patients without vaccine-expanded T cells (non-responders; 13.4 months, $P = 0.003$). Differences in the immune fitness of the patients did not confound this correlation, as responders and non-responders mounted equivalent immunity to a concurrent unrelated mRNA vaccine against SARS-CoV-2. Thus, adjuvant atezolizumab, autogene cevumeran and mFOLFIRINOX induces substantial T cell activity that may correlate with delayed PDAC recurrence.

Main

PDAC is the third leading cause of cancer death in the United States[4] and the seventh worldwide[5]. With an increasing incidence[6], and a survival rate of 12%[1] that has remained largely stagnant for nearly 60 years[1], PDAC is projected to cause even greater global cancer deaths by 2025 (refs. [6,7]).

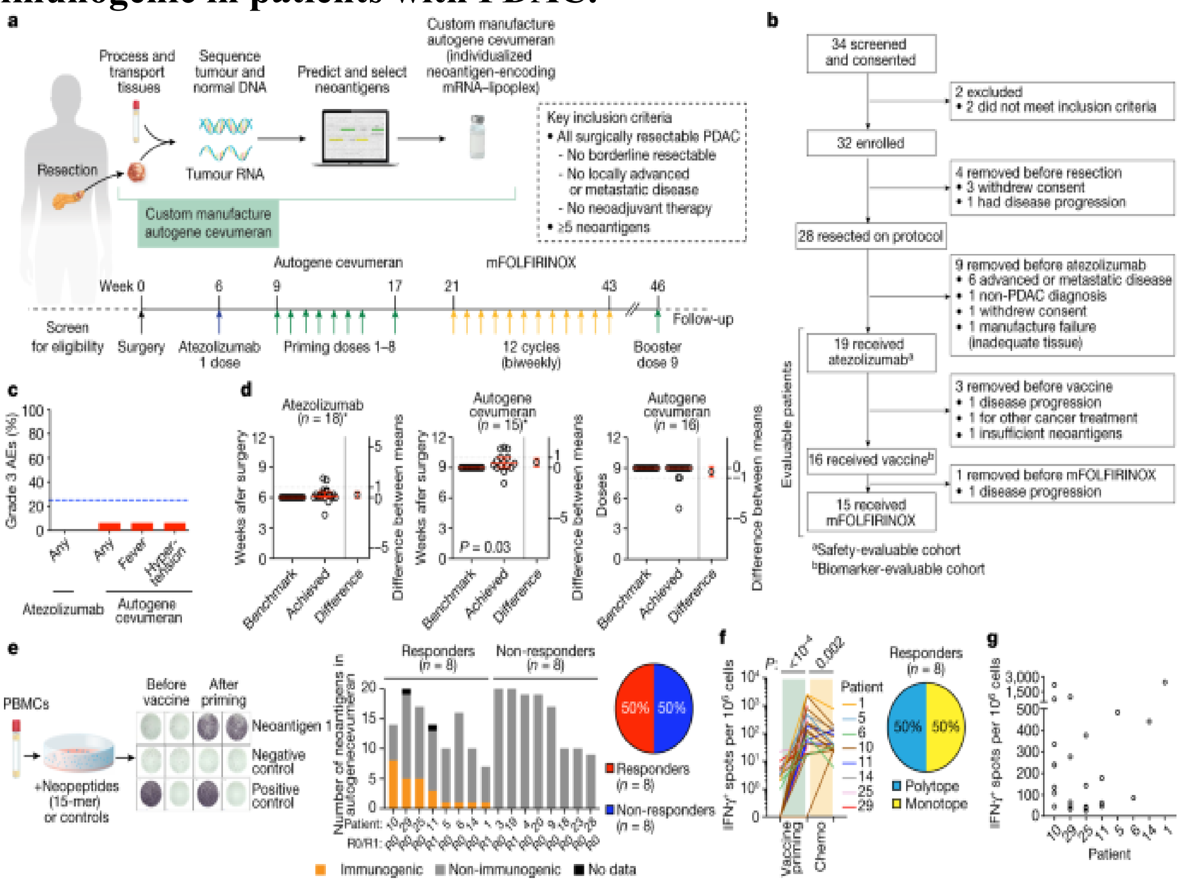
Surgery is the only curative treatment for PDAC. Yet, despite surgery, nearly 90% of patients have disease recurrence at a median of 7–9 months[8,9], and the 5-year overall survival (OS) is only 8–10%[8,9]. Although adjuvant multiagent chemotherapies delay recurrence and are standard of care in surgically resected PDAC, nearly 80% of patients have disease recurrence at around 14 months[4], and their 5-year OS is <30%[10]. Radiation, biologics and targeted therapies are also ineffective[4].

PDACs are almost completely insensitive (<5% response rate[11,12]) to immune checkpoint inhibitors. This insensitivity is partially attributed to the fact that PDACs have a low mutation rate that generates few neoantigens[12], mutation-generated proteins absent from healthy tissues that mark cancers as foreign to T cells, thus potenially rendering PDACs weakly antigenic with few infiltrating T cells. However, recent observations have shown that most PDACs in fact harbour more neoantigens[2,3,13] than previously predicted[14]. Furthermore, studies of long-term survivors of PDAC[2,3] have revealed that neoantigens may stimulate T cells in PDAC. Primary tumours enriched in immunogenic neoantigens also harbour around 12-fold higher densities of activated $CD8^+$ T cells, which correlates with delayed disease recurrence and longer patient survival. Thus, as most PDACs harbour neoantigens with the potential to stimulate T cells, strategies to deliver neoantigens may induce neoantigen-specific T cells and affect patient outcomes.

Based on the observation that long-term survivors of PDAC mount spontaneous T cell responses against tumour-specific neoantigens not shared among patients[2,3], we tested whether adjuvant individualized vaccines can stimulate neoantigen-specific T cells and provide clinical benefit in patients with surgically resected PDAC. Therapeutic mRNA vaccine technology has facilitated the rapid delivery of individualized neoantigen vaccines fully integrated into a routine oncologic workflow[15]. Moreover, mRNA can be rapidly manufactured as individualized vaccines with multiple neoantigens[16], can activate antigen-presenting cells[17,18,19,20] and can be efficiently delivered using newly developed clinical-stage formulations[21]. Therefore, we hypothesized that an effective individualized mRNA vaccine would induce neoantigen-specific T cells in PDAC, eliminate micrometastases and delay recurrence.

To test this hypothesis, we conducted an investigator-initiated, phase I clinical trial of sequential adjuvant atezolizumab (Genentech), autogene cevumeran[22,23] (an individualized mRNA neoantigen vaccine containing up to 20 major histocompatibility complex class I (MHCI) and MHC class II (MHCII) restricted neoantigens in lipoplex nanoparticles intravenously delivered; Individualized NeoAntigen-Specific Therapy (iNeST), BioNTech and Genentech) and mFOLFIRINOX in patients with surgically resectable PDAC (Fig. 1a) to : (1) amplify neoantigen-specific T cells inhibited by PD-1 signalling; and (2) prime naive T cells to vaccine neoantigens.

Fig. 1: Individualized mRNA neoantigen vaccines are safe, feasible and immunogenic in patients with PDAC.



a,**b**, Trial design (**a**) and consolidated standards of reporting trials diagram (**b**). **c**, Percentage of grade 3 AEs attributable to atezolizumab and autogene cevumeran (vaccine) in atezolizumab (*n* = 19) and vaccine (*n* = 16) safety-evaluable patients. Blue line indicates the study-defined safety threshold (25%). **d**, Achieved and benchmarked times to atezolizumab (left) and first vaccine dose (middle), and number of vaccine doses (right). Red line

indicates the median, error bars are 95% confidence intervals and dotted lines the zone of clinical indifference. Asterisks indicate patients on study-specified treatment sequence. **e**–**g**, PBMCs collected after atezolizumab and before vaccine administration, 1–3 weeks after vaccine priming, and 5–6 weeks after mFOLFIRINOX were analysed for IFNγ$^+$ T cells specific to all individual vaccine neoantigens by ex vivo IFNγ ELISpot in $n = 16$ patients in the biomarker-evaluable cohort. Patients were classified as responders if ELISpot detected IFNγ$^+$ T cell induction against at least one vaccine neoantigen. **e**, Left, schematic and representative image of ex vivo IFNγ ELISpot. Middle, Number of vaccine neoantigens per patient that induced IFNγ$^+$ T cells in PBMCs collected after vaccine priming. R0/R1 indicates the surgical margin status. For patient 25, 2 out of 5 ELISpot responses were detected against 2 neoantigen pools (pool 1 with 2 neoantigens, pool 2 with 5 neoantigens). Right, Proportion of vaccine responders and non-responders. **f**,**g**, Normalized ex vivo IFNγ ELISpot counts for vaccine neoantigens that induced a de novo response ($n = 25$ neoantigens in 8 patients): longitudinal (**f**, left); after priming (**g**). Spot counts of the non-stimulated controls were subtracted. Proportion of patients with monotope compared with polytope responses to all vaccine neoantigens (**f**, right). *n* indicates individual patients. Chemo, chemotherapy (mFOLFIRINOX). *P* values calculated using two-tailed unpaired *t*-test (**d**) or Wilcoxon matched-pairs signed-rank test (**f**).

Source Data

Safety, feasibility and immunogenicity

From December 2019 to August 2021, we enrolled 34 patients, of which 28 patients (Fig. 1b) underwent surgery. We then treated 19 patients with atezolizumab, of which 16 patients received subsequent autogene cevumeran. Fifteen out of these 16 patients also received subsequent mFOLFIRINOX (Fig. 1b). We analysed safety in a safety-evaluable cohort ($n = 19$ patients treated with atezolizumab, $n = 16$ treated with autogene cevumeran), and we correlated immune response to RFS in a biomarker-evaluable cohort ($n = 16$ patients treated with atezolizumab and autogene cevumeran). All 19 evaluable participants had clinical characteristics typical of patients with resectable PDAC (Extended Data Fig. 1a). All patients were

treated and followed at the Memorial Sloan Kettering Cancer Center (MSK) during and beyond the enrolment period.

None of the 19 patients treated with atezolizumab in the safety-evaluable cohort had grade 3 or higher adverse events (AEs; Fig. 1c). One out of 16 (6%) patients treated with autogene cevumeran in the safety-evaluable cohort had grade 3 AEs (fever and hypertension; Fig. 1c). All 16 patients (100%) who received autogene cevumeran had grade 1–2 AEs (Extended Data Fig. 1c). We administered atezolizumab and autogene cevumeran at median times within 1 and 3 days of respective benchmarked times (median time to atezolizumab was 6.1 weeks (range of 4.3–7.9 weeks); median time to autogene cevumeran was 9.4 weeks (range of 7.4–11.0 weeks); Fig. 1d). Only 1 patient out of 19 (5%) had insufficient neoantigens that led to non-manufacture of the vaccine (Fig. 1b). Three out of 16 patients (19%) did not receive all 9 vaccine doses (Fig. 1d), which was due to progression, death or mFOLFIRINOX toxicity. Thus, autogene cevumeran can be rapidly administered even after complex oncologic surgery.

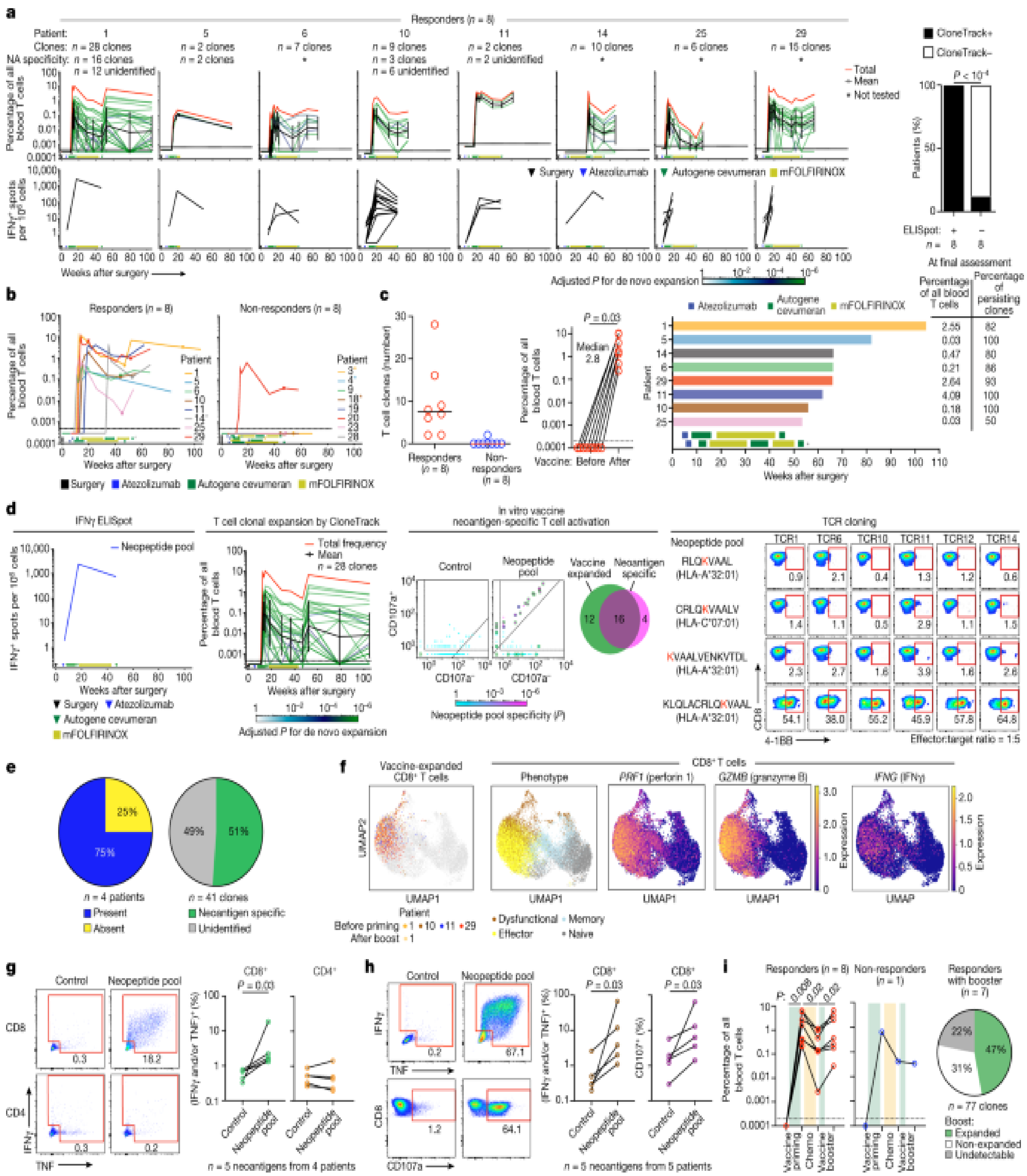
Next, to measure the T cell responses induced by autogene cevumeran, we utilized a previously described ex vivo IFNγ ELISpot assay[24] that detects high-magnitude T cell responses to vaccines without distinguishing $CD8^+$ from $CD4^+$ T cell responses. Eight out of 16 (50%) patients who received the vaccine generated T cell responses that were detected by ex vivo IFNγ ELISpot, and were deemed autogene cevumeran responders (Fig. 1e). By testing each specific target included in the neoantigen vaccines, we detected that 25 out of the 230 neoantigens (11%) administered across all patients who were evaluable at the single-target level induced a T cell response of sufficient high magnitude to be detectable by ex vivo IFNγ ELISpot (Fig. 1e and Extended Data Fig. 2a). Half of all the patients who received the vaccine mounted neoantigen-specific T cell responses against at least one vaccine neoantigen (median = 2, range = 1–8; Extended Data Fig. 2b). Furthermore, half of these responses were polytopic, targeting more than one vaccine neoantigen (Fig. 1e,f). No T cell responses against vaccine neoantigens were detectable before vaccination by ex vivo IFNγ ELISpot (Fig. 1f). Neoantigen-specific immune responses after vaccination were detected at levels that ranged from approximately 100 spots per million bulk peripheral blood mononuclear cells (PBMCs) to >2,000 spots per million bulk PBMCs

(Fig. 1g). Inter-patient variation in the number and magnitude of all responses and intra-patient variation in the magnitude of polytopic responses were observed (Fig. 1g). Thus, autogene cevumeran induces substantial de novo T cell responses in a large proportion of patients with PDAC.

T cell clonotypes and phenotypes

To confirm ELISpot assay reactivity using an orthogonal technique and to probe the diversity and specificity of autogene-cevumeran-expanded T cell clones, we developed CloneTrack. CloneTrack is a new mathematical and immunological method that uses T cell receptor (TCR) Vβ sequencing of peripheral blood samples before and after treatment to identify treatment-expanded high-magnitude T cell clones (Methods). Using CloneTrack, we detected vaccine-induced clonal expansion in 8 out of 8 responders and in 1 out of 8 non-responders (Fig. 2a,b and Extended Data Fig. 2d). In responders, autogene cevumeran expanded multiple clones (median of 7.5 clones; Fig. 2c) from undetectable levels to up to 10% (median of 2.8%; Fig. 2c) of all blood T cells. Analogously, we applied CloneTrack to peripheral blood samples collected before and after atezolizumab treatment and found that autogene-cevumeran-expanded T cell clones did not overlap with atezolizumab-expanded T cell clones (Extended Data Fig. 3a–c). To assess the antigen specificity of autogene-cevumeran-expanded T cell clones, we identified T cell clones specific to ELISpot-identified immunodominant neoantigens in vitro and examined the clonal overlap to autogene-cevumeran-expanded clones in vivo in 2 patients with monotopic responses and in 2 patients with polytopic responses (Fig. 2d and Extended Data Fig. 4a,b). Three out of 4 patients (75%) and 51% ($n = 21$ out of 41 clones) of vaccine-expanded high-magnitude clones (28 out of 41 clones detected in 1 patient) contained immunodominant neoantigen-specific clones (Fig. 2e). In the fourth patient (patient 11), the immunodominant neoantigen-specific clones were in a lower magnitude vaccine-expanded clonal pool (Extended Data Fig. 4b). Thus, autogene cevumeran expands de novo polyclonal neoantigen-specific T cells in PDAC.

Fig. 2: mRNA vaccines expand polyclonal, polyfunctional effector CD8$^+$ T cells.



a, Vaccine-expanded T cell clones assessed using CloneTrack (top), vaccine-induced IFNγ by ELISpot (bottom) and their correlation (right). **b**,**c**, Vaccine-expanded clones identified by CloneTrack: longitudinal aggregate percentage (**b**), number of unique clones (**c**, left), before vaccine and peak expansion aggregate percentage (**c**, middle), and final per patient assessment times (bar graph) with aggregate percentage and clonal fraction at final assessment (**c**, right). For **b**, inverted triangles indicate collection times for

single-cell sequencing in **f** and circles indicate vaccine booster. **d**, Immunodominant vaccine neoantigen-specific clonal overlap with vaccine-expanded clones and specificity to immunodominant vaccine-neoantigens by TCR cloning in patient 1. **e**, Left, Percentage of patients with immunodominant vaccine neoantigen-specific clones in vaccine-expanded clones. Right, Percentage of vaccine-expanded clones specific to immunodominant vaccine neoantigens. **f**, Single-cell phenotypes of vaccine-expanded $CD8^+$ T cells. Dots indicate blood $CD8^+$ T cells. Coloured dots (far left) indicate vaccine-expanded clones in **a**. **g**,**h**, Percentage of $IFN\gamma^+$, TNF^+ and $CD107a^+CD8^+$ (**g**,**h**) and $CD4^+$ T cells (**g**) in PBMCs after vaccine priming with ex vivo immunodominant long (**g**) or minimal (**h**) neopeptide rechallenge. Representative flow plots from patient 1 (**g, h**). Pregated on $CD3^+CD56^-CD8^+$ (**g**,**h**) or $CD4^+$ (**g**) cells. **i**, Left, Aggregate percentage of vaccine-expanded clones with priming, chemotherapy and booster in peripheral blood. Right, Percentage of primed clones that re-expand with booster. *n* indicates individual clones or patients. In **a** and **d**, the green lines indicate individual clone trajectories; the black line, the geometric mean clonal trajectory (error bars are the geometric s.d.); the red line, the cumulative percentage of all expanded clones. In **b,c,** asterisks indicate altered treatment schedules. In **b–d**, rectangles indicate the treatment sequence. *P* values calculated using modified two-tailed Fisher's exact test (**a**,**d**, left), two-sided Chi square test (**a**, right), two-tailed paired *t*-test (**c**), one-tailed binomial test with Bonferroni correction (**d**, middle) or two-tailed Wilcoxon matched-pairs signed-rank test (**g**–**i**).

Source Data

We next studied the phenotype and function of autogene-cevumeran-expanded T cells. Using single-cell RNA sequencing, we found that autogene-cevumeran-expanded high-magnitude clones were $CD8^+$ T cells that expressed lytic markers (perforin 1 and granzyme B) and cytokines (IFNγ) and resembled effector T cells induced by protective viral vaccines[25] (Fig. 2f and Extended Data Fig. 5a–c). Consistently, peripheral blood samples collected after vaccination contained polyfunctional $CD8^+$ T cells but not $CD4^+$ T cells that produced cytokines (IFNγ and TNF) and degranulated on in vitro rechallenge with both long neopeptides (Fig. 2g and Extended Data Fig. 5d) and MHCI-restricted minimal epitopes (Fig. 2h and

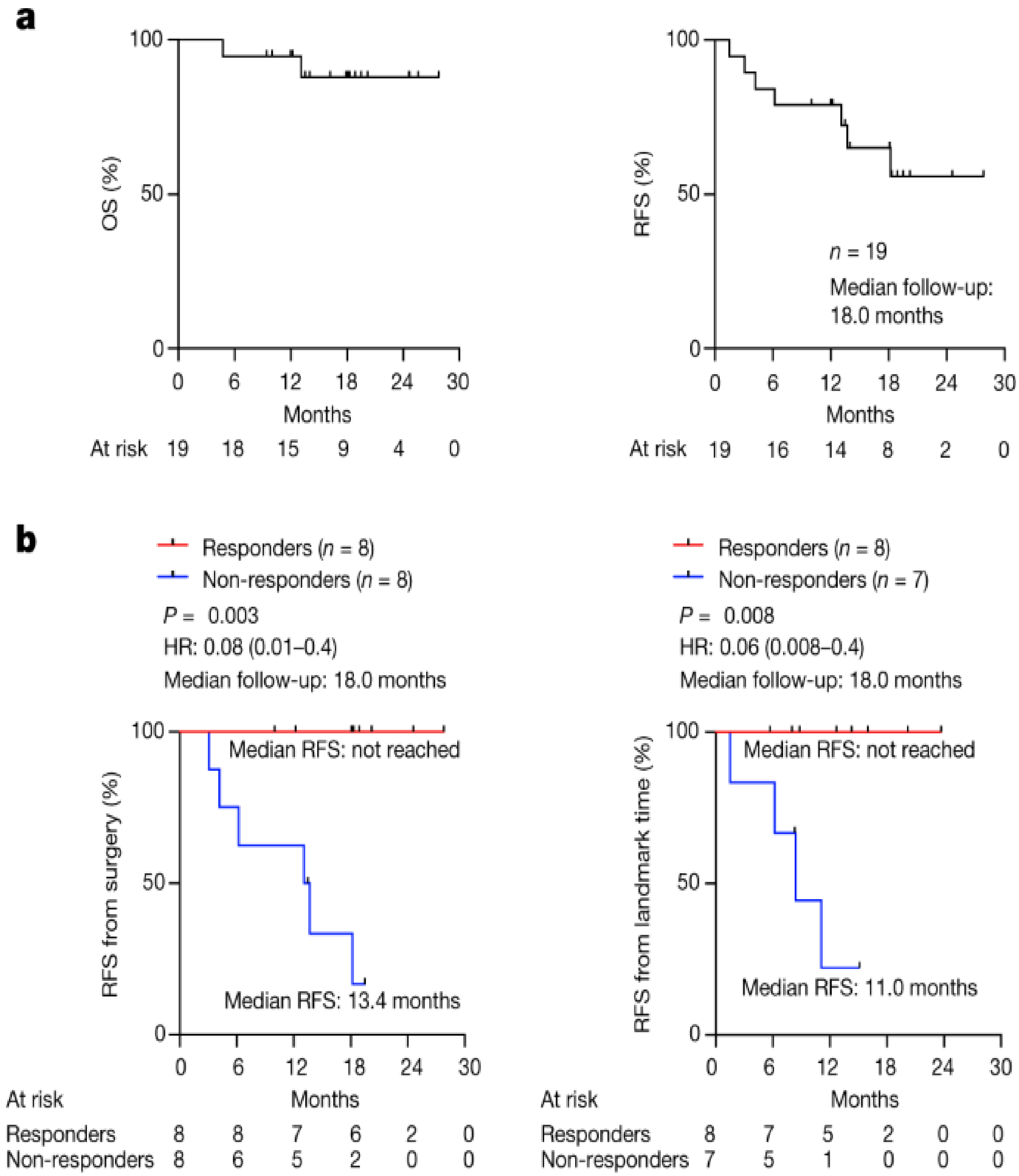
Extended Data Fig. 5e). Notably, autogene-cevumeran-expanded T cells maintained functionality despite post-vaccination mFOLFIRINOX treatment, with persistent IFNγ production (Fig. 1f), uniform re-expansion with a vaccine booster in all responders (Fig. 2i) and sustained persistence as high as 2.5% of all blood T cells up to 2 years after surgery (Fig. 2c and Supplementary Table 1). Furthermore, vaccine boosters re-expanded identical primed clones in 7 out of 7 patients who received boosters (47% of all primed clones; Fig. 2i and Extended Data Fig. 6). Although autogene cevumeran expanded multiple clones, standard flow cytometry did not reliably detect T cell expansion and activation (Extended Data Fig. 5f). Collectively, autogene cevumeran substantially expanded T cells that included vaccine neoantigen-specific, functional and durable $CD8^+$ T cells.

Vaccine response and clinical outcome

At a median follow-up of 18 months that extended beyond the prespecified secondary end point, the median OS and RFS of the patients in the safety-evaluable cohort were not reached (Fig. 3a). For patients in the biomarker-evaluable cohort, the 8 autogene cevumeran responders had a median RFS that was not reached compared with the 8 non-responders who had a median RFS of 13.4 months ($P = 0.003$, hazard ratio (HR) = 0.08 (95% confidence interval (CI) 0.01–0.4); Fig. 3b). To exclude a time-to-response bias[26], we performed a landmark analysis to correlate RFS to response in patients who were recurrence-free when completing all 8 autogene cevumeran priming doses (landmark RFS). The median landmark RFS was similarly not reached in responders compared with 11.0 months in non-responders ($P = 0.008$, HR = 0.06 (95% CI 0.008–0.40); Fig. 3b). Consistently, compared with non-responders, responders had persistently lower serum CA19-9 levels (Extended Data Fig. 7a), the most widely used clinical PDAC biomarker[27]. Only 25% of patients in the biomarker-evaluable cohort had detectable circulating tumour DNA at diagnosis (Extended Data Fig. 7b), as previously reported in patients with resectable PDAC tumours[28,29], and thus was not a reliable biomarker of recurrence. To identify if responders were merely enriched in patients with better prognosis, we found response to atezolizumab, lymph node positivity, margin positivity, primary tumour size, the number of chemotherapy doses and density of intratumoural $CD8^+$ T

cells did not correlate with vaccine response (Extended Data Figs. 1b and 7c–e). Responders and non-responders also had comparable immunological fitness, as they mounted equivalent humoral and cellular responses to an unrelated mRNA vaccine (SARS-CoV-2) that was concurrently administered with autogene cevumeran (Extended Data Fig. 8). Responders and non-responders also had equivalent peripheral frequencies of all major innate and adaptive immune cells (Extended Data Fig. 9a–c), and similar somatic and germline genetic characteristics (Supplementary Tables 2–4). In summary, the autogene-cevumeran-expanded T cell response correlates with delayed PDAC recurrence that is not confounded by detectable differences in patient selection, intratumoural T cell frequency or peripheral T cell frequency or fitness.

Fig. 3: mRNA vaccine response correlates with delayed PDAC recurrence.



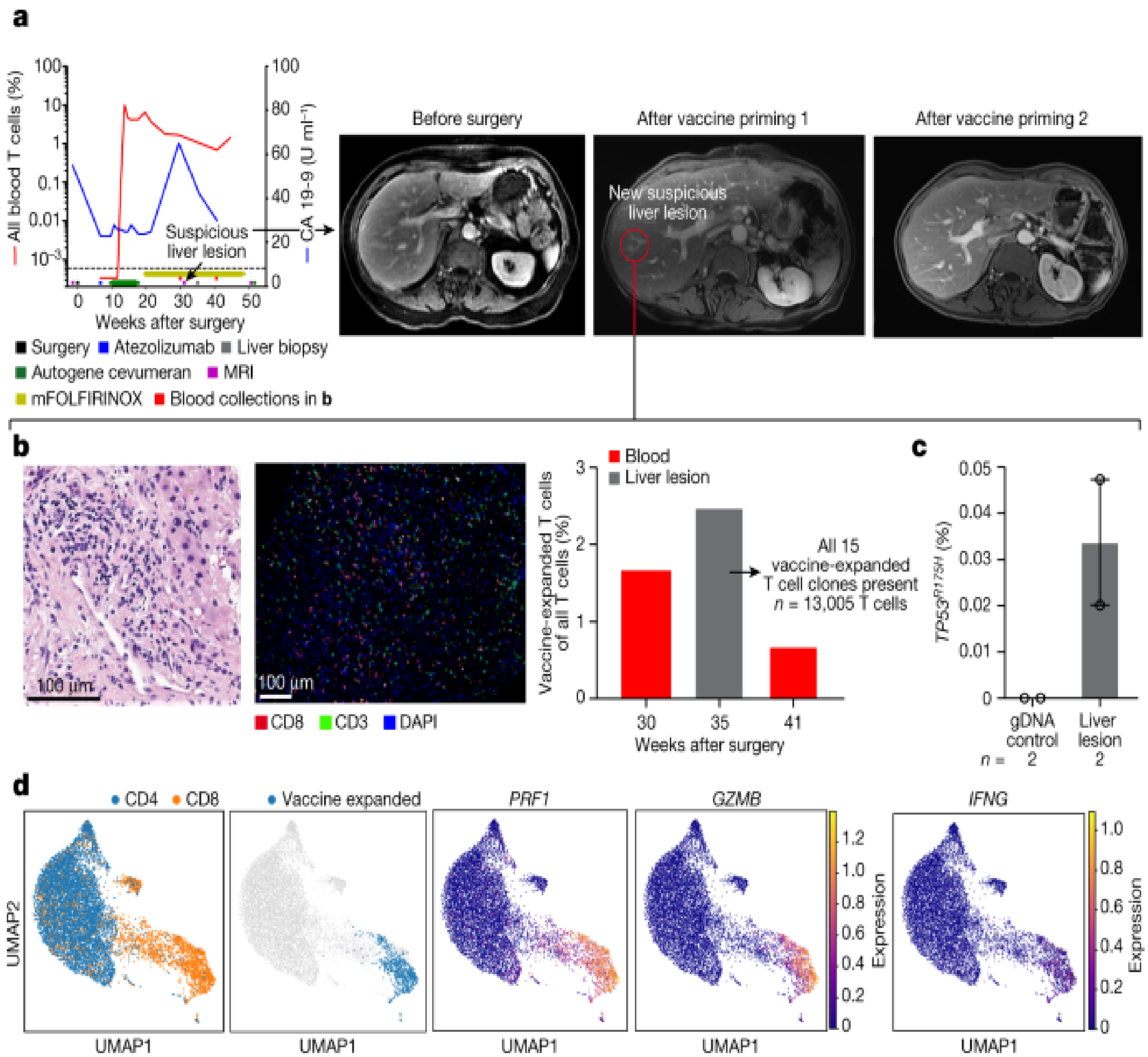
a, OS and RFS in $n = 19$ patients in the safety-evaluable cohort. **b**, RFS from surgery and from landmark time (date of the last vaccine priming dose) stratified by vaccine response in patients in the biomarker-evaluable cohort. *n* indicates individual patients. HR indicates hazard ratio with 95% CI. *P* values calculated using two-tailed log-rank test.

As autogene cevumeran induced high-magnitude T cell responses specific to 25 out of 106 vaccine-encoded neoantigens (24%) in responders (Extended Data Fig. 2a), we searched for correlates of vaccine response. Our previous findings[2,3,30] showed that $CD8^+$ T-cell-enriched PDAC tumours are also enriched in immunogenic 'high-quality' neoantigens distributed in greater proportions across tumour clones. Therefore, we examined whether tumour clonality and neoantigen quality correlate with vaccine-induced T cell responses. Consistently, responders to autogene cevumeran had more clonal tumours than non-responders (Extended Data Fig. 10a). Next, we examined whether immunogenic neoantigens in responders contained high-quality features. We adapted our previously described model[2,3] (Methods) that identifies spontaneously targeted neoantigens in tumours to correlate immunogenicity to the quality of vaccinated neoantigens. In responders, neoantigen quality as a continuous variable correlated with vaccine neoantigens that induced IFNγ ELISpot responses (Extended Data Fig. 10b). Notably, non-responders had similar numbers of non-synonymous mutations and vaccine neoantigens as responders (Extended Data Fig. 2c).

Vaccine clones and a micrometastasis

Patient 29 responded to autogene cevumeran with the second-highest maximal percentage of expanded blood T cells (Fig. 2b) that included vaccine neoantigen-specific polyfunctional $CD8^+$ T cells (Extended Data Fig. 5d,e). Patient 29 developed increased serum CA19-9 levels with a new 7-mm liver lesion suggestive of a metastasis after vaccine priming (Fig. 4a). A biopsy sample did not reveal malignant cells but a dense lymphoid infiltrate (Fig. 4b, left) that included all 15 autogene-cevumeran-expanded (Fig. 4b, middle) $CD8^+$ T cell clones with phenotypic evidence of lytic and effector potential (Fig. 4d). Digital droplet PCR revealed that this lymphoid infiltrate contained rare cells harbouring the *TP53*^{R175H} mutation, identical to the R175H driver mutation in the primary tumour of this patient (Fig. 4c and Extended Data Fig. 10c). This liver lesion disappeared on subsequent imaging (Fig. 4a), which suggests that autogene-cevumeran-expanded T cells may possess the capacity to eradicate micrometastases.

Fig. 4: Vaccine-expanded T cells can infiltrate a micrometastasis.



Clinical and immunological snapshot of a disappearing intrahepatic lymphoid aggregate after vaccination in a patient who responded to the vaccine. **a**, Serial percentage of vaccine-expanded T cells in blood analysed using CloneTrack and serum CA19-9 (left), and abdominal MRI (right) before and after vaccination. **b**, Haematoxylin and eosin staining (left), multiplexed immunofluorescence (middle) and percentage of vaccine-expanded T cells measured using CloneTrack (right, grey bar) in a new liver lesion that developed after vaccination detected by MRI as in **a**. All 15 vaccine-expanded T cell clones (**a**, red line) were present in liver lesion (right, grey bar). **c**, Percentage of mutant *TP53*^{R175H} reads by digital droplet PCR in the liver lesion. The bar indicates the median, the error bars are the s.e.m. **d**, Uniform manifold approximation and projection (UMAP) plots of single-cell phenotypes of all blood T cells (left) and vaccine-expanded

clones (middle), with effector markers (right). *n* indicates the number of T cells detected in liver lesion (**b**) or technical replicates (**c**). Data represent analyses of a single patient.

Source Data

Discussion

We demonstrated that adjuvant autogene cevumeran, an individualized neoantigen vaccine based on uridine mRNA–lipoplex nanoparticles, in combination with atezolizumab and mFOLFIRINOX, is safe, feasible and generates substantial neoantigen-specific T cells in 50% of unselected patients with resectable PDAC. Vaccine-expanded T cells were durable, persisting up to 2 years despite post-vaccination mFOLFIRINOX treatment. High-magnitude vaccine-induced T cell responses, the focus of our immune response analysis that included a new method to track vaccine-expanded clones, correlated with delayed PDAC recurrence. Despite the limited sample size, these early results warrant larger studies of individualized mRNA neoantigen vaccines in PDAC.

As multiple immunotherapies[31] have emerged for immune-inflamed tumours, there remains a need for new immunotherapies for the majority of patients with non-inflamed tumours that are largely insensitive to current immunotherapies. Indeed, the prevailing thought has been that the low passenger mutation rate of such tumours renders them with insufficient neoantigens for vaccines. Here, we provided evidence that despite the low mutation rate of PDAC, a mRNA vaccine can induce T cell activity against neoantigens in this cancer, a non-inflamed tumour with predominantly immune-excluded or desert phenotypes. Whether mRNA neoantigen vaccines can similarly activate T cells in other non-inflamed cancers should be more broadly tested.

We did not find evidence that the correlation of vaccine response to delayed recurrence is confounded by known prognostic variables, such as lymph node or margin-positive disease. Non-responders on average had slightly larger primary tumours than responders; however, larger primary tumour size did not correlate with shorter RFS. As the uridine mRNA–lipoplex

vaccine technology is based on potent antigen delivery into lymphoid compartments and stimulates weak T cell responses in splenectomized mice[20], it is notable that non-responders were also marginally enriched in patients with splenectomies (Extended Data Fig. 1b). Furthermore, that vaccines induced high-magnitude T cell responses in 50% of patients may highlight the need for biomarkers to select optimal patients and tumours for this treatment. Of note, although autogene cevumeran is designed to activate both neoantigen-specific $CD4^+$ and $CD8^+$ T cells and we find it activates high magnitude $CD8^+$ T cells in PDAC, the primary and confirmatory immune response assays in this study do not distinguish $CD8^+$ from $CD4^+$ T cell responses. In fact, as these assays bias towards high-magnitude T cell responses, assays that detect lower magnitude responses may include both $CD4^+$ T cell responses and pre-existing responses. In other tumours, we observed that a substantial proportion of vaccine neoantigens induce de novo responses that are below the ex vivo detectable threshold (manuscript in preparation), a level of response not assessed in this trial. Nevertheless, our results in PDAC imply that a high-magnitude T cell response may contribute to a favourable clinical outcome. Thus, strategies to ensure high-magnitude responses are being pursued, including further optimization of mRNA vaccine potency and extension of the neoantigen discovery space to include genetic aberrations other than single nucleotide variations (SNVs) and insertions and deletions (indels) (for example, fusions)[32]. Notwithstanding, as vaccines expand polyclonal T cells, whether vaccine-induced clonal diversity contributes to durable control[33,34] is another key query for future work.

Our study was not powered to detect differences in biomarkers of vaccine response. Despite this limitation, we observed that tumours in responders were more clonal—possibly representing tumours in immune-edited evolution as seen in immunogenic PDACs in long-term survivors[3]. Thus, we speculate that a more clonal primary tumour may reflect the ability of the immune system to recognize a tumour and therefore respond to a vaccine. Moreover, the observation that neoantigen quality[2,3,30] correlates with immunogenic vaccine neoantigens provides further support for the concept that select neoantigens may possess higher immunogenic qualities that are possibly desirable for vaccines. However, in this trial, as responders and non-responders had comparable numbers of vaccine neoantigens drawn from

a similar number of tumour mutations, we consider that an absence of a response in non-responders is unlikely due to a failure to include immunogenic neoantigens. Overall, these observations remain preliminary but support future investigation of whether tumour clonality and neoantigen quality could serve as biomarkers of vaccine response.

We tested individualized mRNA cancer vaccines in the adjuvant setting motivated by observations that vaccines against pathogens have historically been most effective in preventive and not therapeutic settings, which likely reflects that vaccine efficacy requires an optimally functioning host immune system. In patients with advanced cancer, global impairments in host immunity and knowledge gaps on neoantigen heterogeneity between tumours may hamper neoantigen vaccination. Thus, we propose that vaccines should be tested in patients with minimal residual disease, as is currently ongoing in trials in high-risk colorectal cancer (ClinicalTrials.gov identifier NCT04486378) and in triple-negative breast cancer (ClinicalTrials.gov identifier NCT02316457). Notably, our study demonstrated that mRNA neoantigen vaccines can be individualized in 9 weeks and fully integrated into a standard clinical workflow even after complex oncologic surgery. Given this trial was the index experience with individualized mRNA vaccination for PDAC, mFOLFIRINOX was administered >12 weeks[35,36,37] after surgery. Furthermore, given its limited sample size, this trial enrolled only white individuals. Future studies must test individualized mRNA neoantigen vaccines in a diverse population of patients with PDAC, coupled with a faster time to adjuvant mFOLFIRINOX. Experience with individualized cancer vaccines[16] that predated and accelerated mRNA-based SARS-CoV-2 pandemic vaccines[38] can now further hasten individualized cancer vaccine manufacture times[22,23] and enable more rapid adjuvant custom vaccination and chemotherapy.

Overall, we reported preliminary evidence that adjuvant autogene cevumeran, an individualized mRNA neoantigen vaccine, in combination with atezolizumab and mFOLFIRINOX induces substantial T cell activity in patients with surgically resected PDAC that correlates with delayed recurrence. A follow-up global randomized trial (IMCODE 003, BNT122) is imminent.

Methods

Trial design, treatments, oversight and conduct

We administered atezolizumab, autogene cevumeran and mFOLFIRINOX sequentially to measure how each immunotherapy modulated neoantigen-specific T cells. To establish clinical feasibility, we set the following benchmarked times to treatment after surgery (Fig. 1a): (1) one 1,200 mg intravenous dose of atezolizumab on week 6; (2) nine 25 µg intravenous doses of autogene cevumeran given as seven weekly priming doses beginning on week 9, an eighth dose at week 17 and a ninth booster dose at week 46; (3) 12 cycles of mFOLFIRINOX beginning on week 21. As the half-life of atezolizumab is 27 days, with receptor occupancy persisting for several months[39,40], we hypothesized that this dosing scheme would allow sufficient PD-L1 receptor occupancy to support proficient T cell activation by autogene cevumeran. Additional details are in provided in the protocol in Supplementary Data 1. As we treated the first patient before the COVID-19 pandemic, patients received SARS-CoV-2 vaccines as they became available either before, interspersed during or following completion of individual experimental treatments.

We conducted the study in accordance with the Declaration of Helsinki and good clinical practice guidelines. The study was approved by the institutional review board at MSK, the US Food Drug Administration and was registered on ClinicalTrials.gov (NCT04161755). All participants provided written informed consent.

Immune response assays

We investigated peripheral blood samples using two independent assays. Both, per design, detect high-magnitude T cell responses to vaccines without distinguishing $CD8^+$ from $CD4^+$ T cell responses. IFNγ ELISpot was performed ex vivo (that is, without previous expansion in culture to capture T cells below the assay threshold) to map the induction of neoantigen-specific T cell responses for each vaccine target used across all patients. Patients were classified as responders to autogene cevumeran if the IFNγ

ELISpot assay detected T cell reactivity against at least one vaccine neoantigen.

To confirm IFNγ ELISpot results using an orthogonal technique, we used TCR Vβ sequencing-based CloneTrack to detect greater than twofold in vivo expansion of T cell clones to vaccines in a non-antigen-specific manner. Clones that expanded at different treatment times (before and after atezolizumab; before and after autogene cevumeran) further distinguished atezolizumab from autogene-cevumeran-expanded T cell clones. For 6 ELISpot-identified immunodominant neoantigens in 4 out of 8 (50%) autogene cevumeran responders, we used TCR Vβ sequencing of in vitro neopeptide-stimulated T cells to validate the specificity of in vivo vaccine-expanded T cell clones.

Patients

We enrolled patients with an Eastern Cooperative Oncology Group performance status of 0–1 with single, radiographically suspicious, surgically resectable PDACs, no distant metastases and ≥5 neoantigens. We excluded patients with metastatic, borderline or locally unresectable PDACs, and patients who received neoadjuvant therapy.

After surgery, we included patients with pathologically confirmed PDAC with R0/R1 margins. Additional eligibility criteria and ethical study conduct information are in the protocol (Supplementary Data 1). We aimed to accrue 20 evaluable patients.

Surgery

Patients underwent open pancreaticoduodenectomy or either open or laparoscopic distal pancreatectomy and splenectomy at MSK. We then transported tumour blocks with the most (minimum ≥10%) histological tumour content with matched blood to BioNTech.

End points

The primary end point was safety (Extended Data Table 1). Secondary end points were 18-month RFS and 18-month OS. We defined recurrence as new lesions on the basis of response evaluation criteria in solid tumours (v.1.1), and RFS from either the date of surgery (RFS) or from the date of the last autogene cevumeran priming dose (landmark RFS) to the date of recurrence or death, whichever occurred first. We censored patients without events at the last known date they were recurrence-free. We defined OS from the date of surgery to the date of death. As exploratory end points, we measured immune response and feasibility as actual compared with benchmarked treatment times. Data cut-off was 1 April 2022, extending the median follow-up beyond the prespecified 18-month RFS secondary end point.

Mutation identification and neoantigen selection

DNA was extracted from PBMCs. DNA and RNA were extracted from tumours. Expressed non-synonymous mutations and HLA type were identified by whole-exome sequencing of patient-specific tumour–normal pairs and tumour RNA sequencing. Neoantigens were bioinformatically predicted and ranked by immunogenicity as previously described[16]. mRNA vaccine neoantigen characteristics are detailed in Supplementary Table 5.

Somatic and germline mutation testing

For the detection of somatic tumour mutations in key cancer genes, we used MSK-IMPACT, a previously published targeted tumour-sequencing test that covers 468 cancer genes[41].

For the detection of germline mutations, we utilized the MSK-IMPACT panel to detect pathogenic germline variants. In brief, gDNA was enriched for targeted regions using a hybridization-based protocol and sequenced on an Illumina HiSeq instrument. Sequence reads were aligned to the GRCh37/hg19 reference human genome and variants (SNVs, small indels, and copy number variants encompassing one or more exons) were called using publicly available and in-house developed bioinformatics tools. Variants were classified according to the American College of Medical Genetics and Genomics guidelines[42]. Only variants classified as pathogenic

or likely pathogenic are reported. The list of mutations analysed in the MSK-IMPACT germline panel are provided in Supplementary Table 4.

Production benchmarks

We set the following a priori benchmarks from surgery to vaccine manufacture: (1) transport specimen from operating room to pathology in ≤5 min; (2) fix specimen in formalin and embed in paraffin in ≤15 min; (3) select blocks for vaccine production in ≤2 days; (4) ship to BioNTech in ≤72 h; (5) produce vaccines in ≤6 weeks; (6) administer first dose of the vaccine in ≤9 weeks.

Autogene cevumeran manufacturing

For every patient, individualized mRNA neoantigen vaccines were manufactured under good manufacturing practice conditions containing two uridine-based mRNA strands with noncoding sequences optimized for superior translational performance[43,44]. Each strand encoded up to 10 MHCI and MHCII neoepitopes, formulated in approximately 400 nm diameter lipoplex nanoparticles[20] comprising the synthetic cationic lipid (*R*)-*N*,*N*,*N*-trimethyl-2,3-dioleyloxy-1-propanaminium chloride (DOTMA) and the phospholipid 1,2-dioleoyl-*sn*-glycero-3-phosphatidylethanolamine (DOPE) to enable intravenous delivery.

Cell culture

We purified patient PBMCs from blood samples by density centrifugation over Ficoll-Paque Plus (GE Healthcare). We purified healthy donor PBMCs from buffy coats (New York Blood Center) and isolated T cells using a Pan-T cell isolation kit (Miltenyi Biotech). We activated T cells with CD3/CD28 beads (Thermo Fisher) with IL-7 (3,000 IU ml^{-1}) and IL-15 (100 IU ml^{-1}) (Miltenyi Biotec), and transduced T cells on day 2 after activation. Virus-producing cell lines (H29 and RD114-envelope producers) were as previously described[45,46]. We cultured T cells and K562 cells in RPMI medium supplemented with 10% FBS (Nucleus Biologics), 100 U ml^{-1} penicillin–streptomycin (Thermo Fisher Gibco) and 2 mM l-glutamine (Thermo Fisher Gibco). We cultured patient PBMCs with RPMI medium

supplemented with 10% FBS, 1 mM sodium pyruvate, 2 mM l-glutamine, non-essential amino acids and 2-mercaptoethanol (MSK medium preparation core facility). We cultured H29, RD114-envelope producers and Phoenix-AMPHO in DMEM medium supplemented with 10% FBS (Nucleus Biologics), 100 U ml^{-1} penicillin–streptomycin (Thermo Fisher Gibco) and 2 mM l-glutamine (Thermo Fisher Gibco).

Immune response

Ex vivo IFNγ ELISpot

Multiscreen filter plates (Merck Millipore), precoated with antibodies specific for IFNγ (Mabtech), were washed with PBS and blocked with X-VIVO 15 (Lonza) containing 2% human serum albumin (CSL-Behring) for 1–5 h. Next, 3×10^5 effector cells per well were stimulated for 16–20 h with pools of 15-amino-acid-long peptides (JPT Peptide Technologies) overlapping by 11 amino acids covering the length of each target. Cryopreserved PBMCs were subjected to ELISpot after a resting period of 2–5 h at 37 °C. All tests were performed in duplicate and included anti-CD3 (Mabtech) as a positive control. Bound IFNγ was visualized using a secondary antibody directly conjugated with alkaline phosphatase (ELISpotPro kit, Mabtech). Next, plates were incubated with BCIP/NBT (5-bromo-4-chloro-3′-indolyl phosphate and nitro blue tetrazolium) substrate (ELISpotPro kit, Mabtech). Plates were scanned using an AID Classic Robot ELISPOT reader and analysed using AID ELISPOT 7.0 software (AID Autoimmun Diagnostika). A sample was deemed positive if the IFNγ ELISpot count exceeded a minimum threshold of 7 spots per 300,000 PBMCs. A post-vaccination PBMC sample was deemed positive on the basis of a significant increase in ELISpot count compared with a negative control (medium alone, as no IFNγ ELISpot responses were detected in pre-vaccination samples). To account for varying sample quality reflected in the number of spots in response to anti-CD3 antibody stimulation, we applied a normalization method that enabled direct comparison of spot counts and strength of response between individuals, as described previously[24,47,48]. Statistical significance was determined based on two statistical tests (distribution-free resampling)[47,48].

TCR Vβ clone tracking (CloneTrack)

For TCR Vβ sequencing, we prepared gDNA from bulk PBMCs or purified T cells using a Qiagen DNA extraction kit according to the manufacturer's instructions. We quantified samples using Dropsense 96 and diluted to standard concentrations for library preparation. We generated sample data using an immunoSEQ Assay (Adaptive Biotechnologies). In brief, the somatically rearranged TCRB CDR3 was amplified[49,50] from gDNA using a two-step, amplification bias-controlled multiplex PCR approach. The first PCR consists of forward and reverse amplification primers specific for every known V and J gene segment, and amplifies the hypervariable CDR3 of the immune receptor locus. The second PCR adds a proprietary barcode sequence and Illumina adapter sequences[51]. In addition, reference gene primers were included in the PCR reaction to quantify total nucleated cells that can be sequenced and to accurately measure the fraction of T cells in each sample. CDR3 and reference gene libraries were sequenced on an Illumina instrument according to the manufacturer's instructions. Raw sequence reads were demultiplexed according to Adaptive's proprietary barcode sequences. Demultiplexed reads were further processed to remove adapter and primer sequences and to identify and remove primer dimer, germline and other contaminant sequences. The filtered data were clustered using both the relative frequency ratio between similar clones and a modified nearest-neighbour algorithm to merge closely related sequences to correct for technical errors introduced through PCR and sequencing. The resulting sequences were sufficient to annotate the V, D and J genes and the N1 and N2 regions constituting each unique CDR3 and the translation of the encoded CDR3 amino acid sequence. Gene definitions were based on annotation in accordance with the IMGT database (https://www.imgt.org). The set of observed biological TCRB CDR3 sequences were normalized to correct for residual multiplex PCR amplification bias and quantified against a set of synthetic TCRB CDR3 sequence analogues[50].

T cell clone definition

We identified and tracked T cell clones by their β chain sequence (TRB), defined as the nucleotide CDR3 sequence (including the conserved C and F residues) and a deterministic V and J gene alignment. For T cells identified

by single-cell sequencing, we similarly defined clones by the TRB sequence to map clones to paired TCR Vβ sequencing. Owing to the higher entropy of the nucleotide CDR3 sequence generation probability distribution[52], we used nucleotide instead of amino acid CDR3 sequences to minimize the chance of conflating two different T cell clones (different original VDJ recombination events). This becomes crucial to differentiate clones that may have different α chain (TRA) sequences, which are unobserved in the bulk TCR Vβ sequencing step. We used the provided deterministic V and J alignments from Adaptive Biotechnologies (for bulk TCR Vβ sequencing) and 10x (for single-cell sequencing).

T cell clone frequency estimation

For a given sample of bulk TRB sequences, we estimated the total number of effective cells sequenced, *N*, as the summation of all productive (in-frame, no stop codons) reads. We excluded non-productive reads, as they must necessarily be recombined CDR3s from silenced alleles (we did not model the fraction of productive reads from silenced alleles and assumed them to be a small correction). We estimated a T cell clone x's cell count ${n}_{x}$ in a sample as the number of reads corresponding to the clone as defined above (V and J gene and nucleotide CDR3 sequence). We therefore estimated the frequency of clone x as ${f}_{x}=\frac{{n}_{x}}{N}$. For the purpose of visualization, if we did not observe a clone in a sample, we used a pseudo-frequency of $\frac{1}{3N}$ (if plotting a trajectory with multiple samples, we used the largest *N* over the samples) and indicated this observation threshold as dotted black lines. We computed the aggregate frequency of several clones $x\in X$ in a similar fashion using an aggregate count ${n}_{X}={\sum }_{x\in X}{n}_{x}$ and used the same convention for a pseudo-frequency if ${n}_{X}=0$.

T cell clone significance determination

We took a statistically conservative approach to minimize the false-positive identification of expanded T cell clones. To this end, to calculate treatment-expanded T cell clones, we used a significance threshold of $P_{\text{adjusted}} < 0.001$, where the *P* value is adjusted using Bonferroni correction ($P_{\text{adjusted}} = P \times$ no.

of T cell clones) to account for the large number of T cell clones that were screened.

T cell expansion criteria

To identify treatment-expanded T cell clones, we used an adapted Fisher's exact test and computed *P* values for expanded T cell clones using a two-tailed adapted exact Fisher's test for a twofold increase in a T cell clone between any two samples.

We implemented this by rescaling the repertoire size of the initial sample by half, to effectively reduce the sample size and the number of cells not belonging to the clone in question. We computed this as a Fisher's exact test (implemented from scipy.stats.fisher_exact) using the categorical table supplied in Extended Data Table 2.

Clones that had a fold change <2 (that is, ${\frac{{m}_{x}}{M} < 2\times \frac{{n}_{x}}{N}}$) were assigned a *P* value of 1. These *P* values were then adjusted using Bonferroni correction: ${P}_{{\rm{adjusted}}}=P\times | N\cup M|$, where $|N\cup M|$ designates the number of unique clones in the union of the two samples.

We applied this T cell clone expansion *P* value in the following two contexts to determine whether atezolizumab or autogene cevumeran expanded T cell clones.

Atezolizumab response: to determine whether atezolizumab expanded T cell clones (Extended Data Fig. 3), we compared the number of cells of a particular T cell clone in a blood sample taken on the day of but before atezolizumab administration to the number of cells of that T cell clone in a blood sample taken on the day of but before the first dose of autogene cevumeran. We then considered that a patient had a response to atezolizumab if any T cell clone was found to be significantly expanded ($P_{\rm adjusted} < 0.001$) according to the above outlined expansion criterion.

Autogene cevumeran response with priming doses: to determine whether autogene cevumeran expanded T cell clones, we imposed two criteria: (1) a quality control requirement that a T cell clone must have a minimum of three

reads in at least two samples, and (2) the T cell clone must not be observed before vaccination (0 cells in all samples taken until the day of but before the first dose of autogene cevumeran).

Then, for clones that passed these criteria (as defined above), we compared the number of cells of a particular T cell clone in a blood sample taken after atezolizumab and before vaccination to the number of cells of that T cell clone in any blood sample taken until the day of but before the first dose of mFOLFIRINOX. We further assigned an expansion *P* value as defined as the minimum adjusted expansion *P* value for all samples, further adjusted by Bonferroni correction for the number of samples the expansion *P* values were computed for.

We then considered that a patient had a response to autogene cevumeran if any T cell clone significantly expanded ($P_{adjusted} < 0.001$) according to the above outlined expansion criteria.

For 50% of the responders ($n = 4$ patients), we further examined whether the autogene-cevumeran-expanded T cell clones included neoantigen-specific in vitro clones (see the section ‘In vitro stimulation and T cell clone specificity to peptides’ below).

Autogene cevumeran response with a booster dose: to determine whether an autogene-cevumeran-expanded clone further expanded after the booster dose, we used a standard Fisher’s exact test, with no additional fold change criteria, to compare the clone sizes in samples taken immediately before booster administration to samples from the first follow-up after the booster. As we had previously identified these clones and assessed each one independently, we did not use any multiple hypothesis adjustment and set a significance threshold of $P_{adjusted} < 0.01$ for boost expansion. Boost expansion was analysed in all patients with identified autogene-cevumeran-expanded clones, except for patient 5 who did not receive a booster.

In vitro stimulation and T cell clone specificity to peptides

We resuspended neopeptides (Genscript) in DMSO at 10 mg ml^{-1} and a SARS-CoV-2 peptide pool (Mabtech) following the manufacturer’s

instructions. We stored all peptides at –80 °C. We restimulated PBMCs with peptides in vitro as previously described with minor modifications[2]. In brief, we cultured 1×10^6 PBMCs in a 48-well plate with individual (10 μg ml^{-1}) or pooled (1–5 μg ml^{-1} per peptide) peptides on day 1. We added IL-2 (100 U ml^{-1}) and IL-15 (10 ng ml^{-1}) on day 2 and every subsequent 2–3 days. On day 7, we restimulated cells with peptides and added a CD107a antibody (clone H4A3, PE, BD Biosciences) for 1 h at 37 °C. After 1 h, we added a protein transport inhibitor containing monensin (BD Biosciences) and incubated for 4 h at 37 °C. We then stained the cells for additional surface or intracellular markers as per the manufacturer guidelines, and either analysed or purified cells based on CD107a surface expression, or analysed cells based on intracellular cytokine expression.

To determine whether a T cell clone was specifically stimulated by the peptide pool, we sorted and identified T cell clones in $CD107a^-$ and $CD107a^+$ fractions after peptide stimulation as described above. We then determined a peptide-specificity stimulation P value for each T cell clone using a one-tailed binomial test P value (implementing the scipy.stats.binom_test) with a 0.2 threshold (specifically, significance with respect to at least 20% of a clone being $CD107a^+$ as opposed to $CD107a^-$). We adjusted P values using Bonferroni correction and determined significance at a $P_{adjusted} < 0.001$ threshold. We included DMSO as a control to identify nonspecifically stimulated T cell clones. Of all patients tested, only one nonspecific clone was identified (patient 10) as nonspecifically stimulated in DMSO and both screened peptide pools. This clone was therefore excluded as a peptide-specific clone.

HLA cloning and transduction

We cloned the HLA alleles into an SFG γ-retroviral vector[53] and sequence-verified all plasmids (Genewiz). We transfected Phoenix-AMPHO cells with the plasmids using MegaTran 2.0 (OriGene). We collected virus-containing supernatants 48 h after transfection, added Polybrene (EMD Millipore) and spinoculated K562 cells for 2 h at 2,400 r.p.m. at 33 °C. Seventy-two hours after transduction, we sorted HLA^+ K562 cells using an Aria Cell sorter (BD Biosciences).

TCR cloning, transduction and peptide stimulations

We constructed TCR fragments as previously described[54]. In brief, we isolated TRB V-D-J and TRA V-J sequences from purified, sequenced single T cells and fused the TRB V-D-J and TRA V-J sequences to modified mouse constant TRB and TRA chain sequences[54], respectively (gift from A. Gros) to prevent mispairing of transduced TCRs with the endogenous TCRs[55]. In brief, we joined TRB and TRA chains with a furin SGSG P2A linker, cloned the TCR constructs into a SFG γ-retroviral vector[53] and sequence-verified all plasmids (Genewiz). We transfected H29 cells (gpg29 fibroblasts) with retrovirus vectors using calcium phosphate and produced VSV-G pseudo-typed retroviruses[45]. We next used Polybrene (Sigma) and virus-containing supernatants to generate stable RD114-enveloped producer cell lines[46]. We collected and concentrated virus-containing supernatants using a Retro-X Concentrator (Takara). We then coated non-tissue-culture treated 6-well plates with Retronectin (Takara), plated a titrated virus quantity to 3×10^6 activated T cells per well, centrifuged cells for 1 h at room temperature at 300*g* and used transduced T cells between day 7 and 14 after transduction or cryopreserved them for future use. To stimulate TCR-transduced T cells with peptides, we pulsed 5×10^4 (effector:target ratio 1:1) or 2.5×10^5 (effector:target ratio of 1:5) HLA-transduced K562 cells (antigen presenting cells) in a 96-well U-bottom plate for 1 h at 37 °C with the indicated peptides at the indicated concentrations. After 1 h, we washed the peptide by centrifugation and added 5×10^4 TCR or mock (control) transduced T cells per well. We then measured CD137 (also known as 4-1BB) expression on $CD8^+$ mouse TCR^+ T cells 24 h later.

Immunophenotyping by flow cytometry and optical impedance

For flow cytometry, PBMCs from patients were rested overnight at 37 °C and 5% CO_2 before staining[55]. We defined TCR-transduced $CD8^+$ T cells as live, $CD3^+$, $CD8^+$, mouse TCR^+ cells. We stained cells with the following antibodies: from BioLegend, CD62L (clone DREG-56, BV510), CD56 (clone HCD56, BV605), CD4 (clone OKT4, BV650), CD19 (clone HIB19, BV711), FOXP3 (clone 206D, PE), CD3 (clone SK-7, PE-Cy7), CD8 (clone SK1, FITC or Alexa Fluor 700), CD45RA (clone HI100, APC), CD45

(clone 2D1, Alexa Fluor 700), CD39 (clone A1, BV421), LAG-3 (clone 11C3C65, PerCP-Cy5.5), CD366 (clone F38-2E2, APC-Cy7) CD11c (clone S-HCL-3, BV421), HLA-DR (clone L243, BV785), CD14 (clone HCD14, PE), CD11b (clone ICRF44, APC), IFNγ (clone 4S.B3, BV421), mouse TRB (clone H57-597, PE-Cy5), CD137 (clone 4B4-1, PE), HLA-A,B,C (clone W6/23, APC) and a Zombie Red Fixable Viability kit (423110); from BD Biosciences, PD-1 (clone EH12.1, BV786), TNF (clone MAb11, APC), CD107a (clone H4A3, PE), CD56 (clone NCAM16.2, BV786) and DAPI (564907); from ThermoFisher Scientific, Ki-67 (clone SolA15, PE-Cy5). We stained cells using antibody cocktails in the dark at 4 °C, washed and analysed on a FACS LSR Fortessa (BD Biosciences) using FACSDiva (v.8.0.1) software (BD Biosciences). To examine the expression levels of intracellular markers, we surface-stained, fixed, permeabilized and stained the cells for intracellular proteins using a Fixation and Permeabilization Buffer kit per the manufacturer's recommendations (Invitrogen). We used appropriate FMO controls as indicated. We analysed the data using FlowJo (v.10, Tree Star). We used the following definitions (all pre-gated on live, $CD45^+$ cells): regulatory T cells, $CD3^+CD56^-CD8^-CD4^+FOXP3^+$; dendritic cells, $CD3^-CD56^-CD19^-CD14^-CD11c^+HLA^-DR^+$; monocytes, $CD3^-CD56^-CD19^-CD11b^+CD14^+$; natural killer cells, $CD3^-CD56^+$; B cells, $CD3^-CD19^+$; $CD8^+$ T cells, $CD3^+CD56^-CD8^+CD4^-$; $CD4^+$ T cells, $CD3^+CD56^-CD8^-CD4^+$; natural killer T cells, $CD3^+CD56^+$. To identify frequencies of peripheral blood neutrophils, eosinophils and basophils, we measured respective cell frequencies by optical impedance on a clinical-grade Sysmex analyser.

T cell sorting

We sorted bulk T cells from patient PBMC samples immediately after thawing on a BD FACS Aria flow cytometer (BD Biosciences). We sorted $CD107a^-$ and $CD107a^+$ $CD8^+$ T cells after 7 days of peptide stimulation. We used the sorted T cell samples for TCR Vβ sequencing, single-cell RNA/TCR sequencing, or single-cell TCR sequencing as indicated.

Single-cell RNA/TCR sequencing

Library preparations for single-cell immune profiling, sequencing and post-processing of the raw data were performed at the Epigenomics Core at Weill Cornell Medicine.

Sample preparation

Single-cell RNA sequencing libraries were prepared according to 10x Genomics specifications (Chromium Single Cell V(D)J User Guide PN-1000006, 10x Genomics). Each cellular suspension (>90% viability), at a concentration between 700 and 1,000 cells per µl, was loaded onto to a 10x Genomics Chromium platform to generate Gel Beads-in-Emulsion (GEM), targeting about 10,000 single cells per sample. After GEM generation, polyA cDNA barcoded at the 5′ end by the addition of a template switch oligonucleotide (TSO) linked to a cell barcode and unique molecular identifiers (UMIs) was generated by incubation at 53 °C for 45 min in a C1000 Touch Thermal cycler with a 96-Deep Well Reaction module (Bio-Rad). GEMs were broken and the single-strand cDNA was cleaned up using DynaBeads MyOne Silane Beads (Thermo Fisher Scientific). The cDNA was amplified for 13 cycles (98 °C for 45 s; 98 °C for 20 s, 67 °C for 30 s, 72 °C for 1 h). Quality and quantity of the cDNA was assessed using an Agilent Bioanalyzer 2100, obtaining a product of about 1,600 bp. For generation of 5P expression libraries, an aliquot of the cDNA (about 50 ng) was enzymatically fragmented, end repaired, A-tailed, subjected to a double-sided size selection with SPRI select beads (Beckman Coulter) and ligated to adaptors provided in the kit. A unique sample index for each library was introduced through 14 cycles of PCR amplification using the indexes provided in the kit (98 °C for 45 s; 98 °C for 20 s, 54 °C for 30 s, and 72 °C for 20 s × 14 cycles; 72 °C for 1 min; held at 4 °C). Indexed libraries were subjected to a second double-sided size selection, and libraries were then quantified using Qubit fluorometric quantification (Thermo Fisher Scientific). The quality was assessed on an Agilent Bioanalyzer 2100, obtaining an average library size of 430 bp. For generation of full-length TCR VDJ regions, an aliquot of the cDNA (about 5 ng) was subjected to nested PCR amplification with specific VDJ outer and inner primer pairs (98 °C for 45 s; 98 °C for 20 s, 67 °C for 30 s, and 72 °C for 20 s × 8 cycles; 72 °C for 1 min; held at 4 °C), and one-sided size selection using SPRI select

beads. Quality and quantity of the VDJ region was assessed using an Agilent Bioanalyzer 2100. The average library size was 620 bp.

Sequencing and post-processing of data

5P expression and TCR libraries were clustered on an Illumina NovaSeq pair-end read flow cell and sequenced for 28 cycles on R1 (10x barcode and the UMIs), followed by 8 cycles of I7 Index (sample Index), and 91 bases on R2 (transcript), obtaining about 250 million clusters for 5P expression and 50 million for TCR libraries. Primary processing of sequencing images was done using Illumina's Real Time Analysis software (RTA). 10x Genomics Cell Ranger Single Cell Software suite (https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger was used to perform sample) was used for demultiplexing, alignment (hg19), filtering, UMI counting, single-cell 5′ end gene counting, TCR assembly, annotation of paired VDJ and performing quality control using the manufacturer's parameters.

Analysis

Filtered gene expression matrices generated from 10x CellRanger for five samples were matched to paired TCR sequences using the python package Scirpy[56]. All five samples were aggregated into a single unnormalized counts matrix and all downstream analyses were performed using GeneVector[57]. Batch-effect correction was applied over all cells using the samples as batch labels. Cells were first classified as either CD4 or CD8 T cells using the respective gene marker. $CD8^+$ T cells were further classified into four phenotypes (effector, memory, naive and dysfunctional) using previously published gene markers[25,58]. A probability distribution over phenotypes was generated for each cell, and phenotype assignment corresponded to the maximum probability. Vaccine-specific T cells were identified by exact match of the associated nucleotide sequence. Uniform manifold approximation and projection visualizations were constructed using the python library Scanpy[59].

Clonality

Whole-exome sequence reads of tumour–normal paired samples of patients were aligned to the reference human genome (hg19) using the Burrows–Wheeler alignment tool (bwa mem v.0.7.17) and samtools (v.1.6). Duplicates were marked with picard-2.11.0 MarkDuplicates (http://broadinstitute.github.io/picard). Indel realignments were done using the Genome Analysis toolkit (GenomeAnalysisTK-3.8-1-0-gf15c1c3ef) RealignerTargetCreator and IndelRealigner[60] using 1000 genome phase1 indel (1000G_phase1.indels.b37.vcf) and Mills indel calls (Mills_and_1000G_gold_standard.indels.b37.vcf) as references. Base calls were recalibrated using BaseRecalibrator and dbSNP (v.138). Both tumour samples were covered at 378× and normal samples at 346× on average on its target regions.

MuTect 1.1.7 and Strelka 1.0.15 were used to call SNV and indels on pre-processed sequencing data. For the MuTect calls, dbSNP 138 and CosmicCodingMuts.vcf (v.86)[61] were used as reference files. For the Strelka calls, we set "isSkipDepthFilters = 1" to prevent filtering-out of mutation calls from exome sequencing due to exome-sequencing mapping breadth. Unbiased normal and tumour read counts for each SNV and indel call were then assigned with the bam-readcount software 0.8.0-unstable-6-963acab-dirty (commit 963acab-dirty) (https://github.com/genome/bam-readcount). A minimum base quality filter was set with the "-b 15" flag. The reads were counted in an insertion-centric way with the "-i" flag, so that reads overlapping with insertions were not included in the per-base read counts. We then used the normal and tumour read counts to filter mutations. The following filtering criteria were used: (1) total coverage for tumour ≥10; (2) variant allele frequency for tumour ≥2%; (3) number of reads with alternative allele ≥5 for tumour; (4) total coverage for normal ≥7; and (5) variant allele frequency for normal ≤1% at a given mutation. Filtered mutation sets were annotated using SnpEff (v.4.3t). 23 Dbsnp138 (b37) was used for snp-pileup.

To infer clonality of vaccine targets, we extracted missense and frameshift mutations from the filtered VCF files, and these mutations were put into the PhyloWGS software package (https://github.com/morrislab/phylowgs, v.1.0-

rc2, branch: 681df79) along with copy number variant calls for phylogeny reconstruction. Among 10,000 trees from PhyloWGS, we took the top five trees based on the likelihood and computed the average entropy level to measure tumour heterogeneity. For a given tree, we computed exclusive clone frequencies such as

$${x}^{\alpha }={X}^{\alpha }-\sum _{\beta \in D(\alpha)}{X}^{\beta }$$

where $D(\alpha)$ is the set of clones that are direct descendants of clone α in the given tree, and ${X}^{\alpha }$ is the cellular cancer fraction of clone α. Based on exclusive clone frequencies, we computed Shannon's entropy as a measure of tumour heterogeneity as follows:

$$S={\left\langle -\sum _{\alpha \in \tau }{x}^{\alpha }{\rm{\log }}({x}^{\alpha })\right\rangle }_{\tau }$$

where ${\langle .\rangle }_{\tau }$ is the arithmetic average operator from top five trees (τ).

Neoantigen quality and vaccine response

To model neoantigen quality, we adapted our previously described model[2,3] that identified spontaneously targeted neoantigens in tumours to identify immunogenic neoantigens for vaccination. Specifically, according to our model[2,3] the immunogenicity (or quality) of a neoantigen is the product of two components. The first component—the non-self-recognition potential, *R*, of a neoantigen—is the inherent immunogenicity of the neopeptide. The second component—the self-discrimination potential, *D*—models whether the cognate T cells of a neoantigen avoid negative thymic selection to therefore render neoantigen recognition less constrained by self-toleration.

Previous versions of our quality model[2,3] estimated the non-self-recognition potential *R* of a neopeptide using sequence homology (as determined by soft max rescaling of BLAST alignment) to the immunogenic infectious disease-derived epitopes in the Immune Epitope Database (IEDB). Self-discrimination was estimated as a sum of two free discrimination energies between the neoantigen (MT) and its wild-type (WT) peptide, one for

differential MHC presentation, the other for differential T cell cross reactivity:

$$D=\log \left(\frac{{K}_{{\rm{d}}}^{{\rm{WT}}}}{{K}_{{\rm{d}}}^{{\rm{MT}}}}\right)\,+\,| \log \left(\frac{{{\rm{EC}}}_{50}^{{\rm{MT}}}}{{{\rm{EC}}}_{50}^{{\rm{WT}}}}\right)| $$

where K_d is the HLA specific peptide–MHC affinity (as estimated by netMHC 3.4), and EC_{50} is the concentration for 50% activation for an avidity curve with the neopeptide and its cognate T cell clone[2,3]. Furthermore, in previous studies, we restricted our definition of minimal epitopes to consider to only 9-mers, the most common length of MHCI-bound peptides, predicted to bind to the HLA of the patient with a cut-off of 500 nM.

To now extend the notion of the cross-reactivity or epitope distance beyond the single substitution case as previously described[2,3], we now made an independent site approximation by modelling the cross-reactivity distance, d_C, between two 9-mer epitopes, p^A and p^B as follows:

$${d}_{{\rm{C}}}({p}^{{\rm{A}}},{p}^{{\rm{B}}})=\mathop{\sum }\limits_{i=1}^{9}{d}_{i}{M}_{{p}_{i}^{{\rm{A}}},{p}_{i}^{{\rm{B}}}}\approx | \log \left(\frac{{{\rm{EC}}}_{50}({p}^{{\rm{A}}})}{{{\rm{EC}}}_{50}({p}^{{\rm{B}}})}\right)| $$

where d_i is a scaling weight for position i and M is the substitution matrix as inferred from ref. [3]. This extension allowed us to replace the estimation of the non-self-recognition potential R of a neopeptide from sequence homology using BLAST with epitope distance.

We now took as our two components, in the context of vaccination, how far a neopeptide is from the germline and how close it is to known antigenic IEDB epitopes. For a given 9-mer minimal neoepitope, p^{MT} we defined the quality of the 9-mer as follows:

$$Q({p}^{{\rm{M}}{\rm{T}}})={d}_{{\rm{C}}}({p}^{{\rm{M}}{\rm{T}}}\,,{p}^{{\rm{W}}{\rm{T}}})-\mathop{{\rm{\min }}}\limits_{{p}^{{\rm{a}}}\in {P}_{{\rm{I}}{\rm{E}}{\rm{D}}{\rm{B}}}}{d}_{{\rm{C}}}({p}^{{\rm{M}}{\rm{T}}}\,,{p}^{{\rm{a}}})$$

where P_{IEDB} is the collection of all 9-mers sequences and subsequences of IEDB epitopes.

We defined the quality of a neopeptide as the average quality over the two highest quality 9-mer subsequences that include the substituted residue and are predicted binders (threshold of 4,000 nM) to the HLA type of the individual. As a vaccine can induce neoantigen expression in excess of endogenous expression in a tumour, we dropped the differential MHC presentation term and relaxed our MHC binding cut-off.

To determine whether neoantigen quality correlated with immunogenicity of neoantigenic peptides included in the vaccines used in this study, we classified the neopeptides from the $n = 8$ immune responders as derived from immunogenic or non-immunogenic neoantigens according to the ELISpot assay. Individual immunogenicity was unable to be established for 7 out of the neoantigens from patient 25 and were excluded from the analysis. We used neoantigens only from immune responders to ensure that lack of an immunological response to a neoantigen reflected non-immunogenicity and not general vaccine failure. This generated 23 immunogenic neoantigens out of a total of 99 screened neoantigens from $n = 8$ immune responders. After excluding neoantigens with no predicted minimal epitope binders, we had a final cohort of 22 reactive neopeptides out of a total of 79.

Immunofluorescence

Automated double immunofluorescence was conducted using the Leica Bond BX staining system. Paraffin-embedded tissues were sectioned at 5 μm and baked at 58 °C for 1 h. Slides were loaded in Leica Bond and staining was performed as follows. Samples were pretreated with EDTA-based epitope retrieval ER2 solution (Leica, AR9640) for 20 min at 95 °C. The double antibody staining and detection were conducted sequentially. Primary

antibodies against CD3 (0.6 μg ml^{-1}, rabbit, Dako, A0452) and CD8 (1/10, rabbit, Ventana (Roche), 790-4460) were used. The Leica Bond Polymer anti-rabbit HRP secondary antibody (Leica Biosystems, DS9800) was applied followed by Alexa Fluor tyramide signal amplification reagents (Life Technologies, B40953) or CF dye tyramide conjugates (Biotium, 92174) for detection. After CD3 staining, epitope retrieval was performed for denaturation of primary and secondary antibodies before CD8 antibody was applied. After the run was finished, slides were washed in PBS and incubated in 5 μg ml^{-1} 4′,6-diamidino-2-phenylindole (DAPI) (Sigma Aldrich) in PBS for 5 min, rinsed in PBS and mounted in Mowiol 4–88 (Calbiochem). Slides were kept overnight at −20 °C before imaging. Slides were scanned on a Panoramic scanner (3DHistech) using a ×20/0.8 NA objective. Whole tissues were annotated in CaseViewer (3DHistech) and converted to Tiff images. ImageJ was used to segment cells based on DAPI and to quantify whether a given cell is single, double or null positive.

Humoral responses to SARS-CoV-2

We measured anti-SARS-CoV-2 spike IgG antibody titres with a chemiluminescent microparticle immunoassay (AdviseDx SARS-CoV-2 IgG II assay; Abbott). In brief, we combined serum samples with paramagnetic particles coated with recombinant SARS-CoV-2 protein specific for the receptor-binding domain of the S1 protein, followed by incubation, washing and addition of a conjugate and chemiluminescent substrate. We then measured the resulting chemiluminescent reaction as a relative light unit, with a direct relationship between the amount of IgG antibodies to SARS-CoV-2 in the sample and the relative light unit detected by the system optics (Architect i2000 analyzer). We used a 4 Parameter Logistic Curve fit data reduction method (4PLC, Y-weighted) to generate calibration, with a positivity cut-off of 50.0 AU ml^{-1}.

Circulating tumour DNA

We measured circulating tumour DNA using MSK-ACCESS[62], a high-depth next generation sequencing assay with molecular barcoding technology for the detection of very low frequency somatic alterations in 129 key cancer-associated genes within the plasma cell-free DNA (cfDNA) fraction in

peripheral blood. In brief, cfDNA MSK-ACCESS raw sequence data were demultiplexed and processed as previously described[62]. cfDNA samples were sequenced to a median raw coverage of 25,465× (range 5,007×–49,869×); after collapsing, the median duplex coverage was 1,088× (range 187×–1,783×). Variant calling was performed in a matched tumour-informed manner ("genotyping") using GetBaseCountMultiSample (https://github.com/msk-access/GetBaseCountsMultiSample) and required at least 1 duplex consensus read or 2 simplex consensus reads, to call a somatic SNV or short indel at a site known to be altered in the matched tumour sample from a given patient, as previously described[62].

Digital droplet PCR

DNA extraction

Formalin-fixed paraffin-embedded (FFPE) curls collected in AutoLys M tubes (Thermo Fisher, A38738) were digested with protease solution. DNA was extracted using a MagMAX FFPE DNA/RNA Ultra kit (Thermo Fisher, A31881) on a KingFisher Flex purification system (Thermo Fisher) according to the manufacturer's protocol. Samples were eluted in 55 µl elution solution.

Detection of *TP53* ^{R175H} mutation by digital droplet PCR

TP53 assays were ordered from Bio-Rad (assay identifiers dHsaCP2000105 for TP53 p.R175H c.524G>A; dHsaCP2000106 for TP53 WT). Cycling conditions were tested to ensure optimal annealing and extension temperatures and optimal separation of positive from empty droplets. Optimization was done using a known positive control.

After PicoGreen quantification, 9 ng of gDNA was combined with locus-specific primers, FAM-labelled and HEX-labelled probes, HaeIII and digital PCR Supermix for probes (no dUTP). All reactions were performed on a QX200 ddPCR system (Bio-Rad, 1864001), and each sample was evaluated in technical duplicates. Reactions were partitioned into a median of around 22,000 droplets per well using a QX200 droplet generator. Emulsified PCR assays were run on a 96-well thermal cycler using cycling conditions

identified during the optimization step (95 °C for 10 min; 40 cycles of 94 °C for 30 min and 55 °C for 1 min; 98 °C for 10 min; 4 °C hold). Plates were read and analysed using QuantaSoft software to assess the number of droplets positive for mutant DNA, WT DNA, both or neither.

Statistical analyses

Safety end points are presented descriptively as percentages. Sample sizes (n) represent the number of patients, tumours, T cell clones or neoantigens. We analysed feasibility as the statistical equivalence between benchmarked and achieved treatment times. Here, we defined a delay of <1 week as the zone of clinical indifference and defined the achieved time to be statistically equivalent to the benchmarked time if the 90% confidence interval of the achieved time was within the zone of clinical indifference. We analysed survival curves using log-rank (Mantel Cox) test, compared two groups using unpaired two-tailed Mann–Whitney test and categorical variables using Chi square test. We compared longitudinal clonal expansion using two-tailed Fisher's exact test, and in vitro clonal activation using binomial test with Bonferroni correction. $P < 0.05$ was considered significant. All analyses were performed using GraphPad Prism (v.9.3.1) or Python (v.3.4).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The clinical protocol approved by the institutional review board is provided in Supplementary Data 1. All single-cell sequencing data are available at the Gene Expression Omnibus (accession number GSE222011). De-identified individual participant data reported in the paper will be shared under data use agreements upon reasonable request. Requests must be made to balachav@mskcc.org. Source data are provided with this paper.

Code availability

The following codes are available at GitHub: to infer clonality, https://github.com/genome/bam-readcount, https://github.com/mskcc/facets and https://github.com/morrislab/phylowgs (with details outlined in the Methods); to construct and apply the neoantigen quality model, https://github.com/LukszaLab/NeoantigenEditing; and for CloneTrack, https://github.com/zsethna/CloneTrack.

References

1. Siegel, R. L., Miller, K. D., Wagle, N. S. & Jemal, A. Cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 17–48 (2023).

2. Balachandran, V. P. et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).

3. Łuksza, M. et al. Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389–395 (2022).

4. Park, W., Chawla, A. & O'Reilly, E. M. Pancreatic cancer. *JAMA* **326**, 851–862 (2021).

5. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **71**, 209–249 (2020).

6. Rahib, L. et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74**, 2913–2921 (2014).

7. GBD 2017 Pancreatic Cancer Collaborators The global, regional, and national burden of pancreatic cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol. Hepatol.* **4**, 934–947 (2019).

8. Oettle, H. et al. Adjuvant chemotherapy with gemcitabine and long-term outcomes among patients with resected pancreatic cancer: the

CONKO-001 randomized trial. *JAMA* **310**, 1473–1481 (2013).

9. Neoptolemos, J. P. et al. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. *N. Engl. J. Med.* **350**, 1200–1210 (2004).

10. Strobel, O., Neoptolemos, J., Jäger, D. & Büchler, M. W. Optimizing the outcomes of pancreatic cancer surgery. *Nat. Rev. Clin. Oncol.* **16**, 11–26 (2018).

11. Royal, R. E. et al. Phase 2 trial of single agent Ipilimumab (anti-CTLA-4) for locally advanced or metastatic pancreatic adenocarcinoma. *J. Immunother.* **33**, 828–833 (2010).

12. Yarchoan, M., Hopkins, A. & Jaffee, E. M. Tumor mutational burden and response rate to PD-1 inhibition. *N. Engl. J. Med.* **377**, 2500–2501 (2017).

13. Bailey, P. et al. Exploiting the neoantigen landscape for immunotherapy of pancreatic ductal adenocarcinoma. *Sci. Rep.* **6**, 35848 (2016).

14. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).

15. Vormehr, M., Türeci, Ö. & Sahin, U. Harnessing tumor mutations for truly individualized cancer vaccines. *Annu. Rev. Med.* **70**, 395–407 (2019).

16. Sahin, U. et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).

17. Heil, F. et al. Species-specific recognition of single-stranded RNA via Toll-like receptor 7 and 8. *Science* **303**, 1526–1529 (2004).

18. Diebold, S. S., Kaisho, T., Hemmi, H., Akira, S. & e Sousa, C. R.Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science* **303**, 1529–1531 (2004).

19. Lund, J. M. et al. Recognition of single-stranded RNA viruses by Toll-like receptor 7. *Proc. Natl Acad. Sci. USA* **101**, 5598–5603 (2004).

20. Kranz, L. M. et al. Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* **534**, 396–401 (2016).

21. Sahin, U. et al. An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. *Nature* **585**, 107–112 (2020).

22. Braiteh, F. et al. Abstract CT169: a phase Ia study to evaluate RO7198457, an individualized neoantigen specific immunotherapy (iNeST), in patients with locally advanced or metastatic solid tumors. *Cancer Res.* **80**, CT169 (2020).

23. Lopez, J. S. et al. Abstract CT301: a phase Ib study to evaluate RO7198457, an individualized neoantigen specific immunotherapy (iNeST), in combination with atezolizumab in patients with locally advanced or metastatic solid tumors. *Cancer Res.* **80**, CT301 (2020).

24. Sahin, U. et al. BNT162b2 vaccine induces neutralizing antibodies and poly-specific T cells in humans. *Nature* **595**, 572–577 (2021).

25. Akondy, R. S. et al. Origin and differentiation of human memory CD8 T cells after vaccination. *Nature* **552**, 362–367 (2017).

26. Anderson, J. R., Cain, K. C. & Gelber, R. D. Analysis of survival by tumor response. *J. Clin. Oncol.* **1**, 710–719 (1983).

27. Mizrahi, J. D., Surana, R., Valle, J. W. & Shroff, R. T. Pancreatic cancer. *Lancet* **395**, 2008–2020 (2020).

28. Groot, V. P. et al. Circulating tumor DNA as a clinical test in resected pancreatic cancer. *Clin. Cancer Res.* **25**, 4973–4984 (2019).

29. Hussung, S. et al. Longitudinal analysis of cell-free mutated KRAS and CA 19–9 predicts survival following curative resection of pancreatic cancer. *BMC Cancer* **21**, 49 (2021).

30. Łuksza, M. et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517–520 (2017).

31. Carlino, M. S., Larkin, J. & Long, G. V. Immune checkpoint inhibitors in melanoma. *Lancet* **398**, 1002–1014 (2020).

32. Weber, D. et al. Accurate detection of tumor-specific gene fusions reveals strongly immunogenic personal neo-antigens. *Nat. Biotechnol.* **40**, 1276–1284 (2022).

33. Ott, P. A. et al. A phase Ib trial of personalized neoantigen therapy plus anti-PD-1 in patients with advanced melanoma, non-small cell lung cancer, or bladder cancer. *Cell* **183**, 347–362.e24 (2020).

34. Hu, Z. et al. Personal neoantigen vaccines induce persistent memory T cell responses and epitope spreading in patients with melanoma. *Nat. Med.* **27**, 515–525 (2021).

35. Valle, J. W. et al. Optimal duration and timing of adjuvant chemotherapy after definitive surgery for ductal adenocarcinoma of the pancreas: ongoing lessons from the ESPAC-3 study. *J. Clin. Oncol.* **32**, 504–512 (2014).

36. Mirkin, K. A., Greenleaf, E. K., Hollenbeak, C. S. & Wong, J. Time to the initiation of adjuvant chemotherapy does not impact survival in patients with resected pancreatic cancer. *Cancer* **122**, 2979–2987 (2016).

37. Xia, B. T. et al. Time to initiation of adjuvant chemotherapy in pancreas cancer: a multi-institutional experience. *Ann. Surg. Oncol.* **24**, 2770–2776 (2017).

38. US FDA. Vaccines and Related Biological Products Advisory Committee June 28 2022, Meeting Announcement https://www.fda.gov/advisory-committees/advisory-committee-calendar/vaccines-and-related-biological-products-advisory-committee-june-28-2022-meeting-announcement#event-materials (US FDA, 2022).

39. Herbst, R. S. et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014).

40. Stroh, M. et al. Clinical pharmacokinetics and pharmacodynamics of atezolizumab in metastatic urothelial carcinoma. *Clin. Pharmacol. Ther.* **102**, 305–312 (2017).

41. Cheng, D. T. et al. Memorial Sloan Kettering–integrated mutation profiling of actionable cancer targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. *J. Mol. Diagn.* **17**, 251–264 (2015).

42. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–423 (2015).

43. Holtkamp, S. et al. Modification of antigen-encoding RNA increases stability, translational efficacy, and T-cell stimulatory capacity of dendritic cells. *Blood* **108**, 4009–4017 (2006).

44. Kreiter, S. et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature* **520**, 692–696 (2015).

45. Gallardo, H. F., Tan, C., Ory, D. & Sadelain, M. Recombinant retroviruses pseudotyped with the vesicular stomatitis virus G glycoprotein mediate both stable gene transfer and pseudotransduction in human peripheral blood lymphocytes. *Blood* **90**, 952–957 (1997).

46. Ghani, K. et al. Efficient human hematopoietic cell transduction using RD114- and GALV-pseudotyped retroviral vectors produced in suspension and serum-free media. *Hum. Gene Ther.* **20**, 966–974 (2009).

47. Moodie, Z., Huang, Y., Gu, L., Hural, J. & Self, S. G. Statistical positivity criteria for the analysis of ELISpot assay data in HIV-1 vaccine trials. *J. Immunol. Methods* **315**, 121–132 (2006).

48. Moodie, Z. et al. Response definition criteria for ELISPOT assays revisited. *Cancer Immunol. Immunother.* **59**, 1489–1501 (2010).

49. Robins, H. S. et al. Digital genomic quantification of tumor-infiltrating lymphocytes. *Sci. Transl. Med.* **5**, 214ra169 (2013).

50. Carlson, C. S. et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* **4**, 2680 (2013).

51. Robins, H. et al. Ultra-sensitive detection of rare T cell clones. *J. Immunol. Methods* **375**, 14–19 (2012).

52. Sethna, Z., Elhanati, Y., Callan, C. G., Walczak, A. M. & Mora, T. OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics* **35**, 2974–2981 (2019).

53. Rivière, I., Brose, K. & Mulligan, R. C. Effects of retroviral vector design on expression of human adenosine deaminase in murine bone marrow transplant recipients engrafted with genetically modified cells. *Proc. Natl Acad. Sci. USA* **92**, 6733–6737 (1995).

54. Gros, A. et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* **22**, 433–438 (2016).

55. Cohen, C. J., Zhao, Y., Zheng, Z., Rosenberg, S. A. & Morgan, R. A. Enhanced antitumor activity of murine–human hybrid T-cell receptor (TCR) in human lymphocytes is associated with improved pairing and TCR/CD3 stability. *Cancer Res.* **66**, 8878–8886 (2006).

56. Sturm, G. et al. Scirpy: a Scanpy extension for analyzing single-cell T-cell receptor sequencing data. *Bioinformatics* **36**, 4817–4818 (2020).

57. Ceglia, N. et al. GeneVector: identification of transcriptional programs using dense vector representations defined by mutual information. Preprint at *bioRxiv* https://doi.org/10.1101/2022.04.22.487554 (2022).

58. Leun, A. M., van der Thommen, D. S. & Schumacher, T. N. $CD8^+$ T cell states in human cancer: insights from single-cell analysis. *Nat. Rev. Cancer* **20**, 218–232 (2020).

59. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

60. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

61. Tate, J. G. et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2018).

62. Rose Brannon, A. et al. Enhanced specificity of clinical high-sensitivity tumor mutation profiling in cell-free DNA via paired normal sequencing using MSK-ACCESS. *Nat. Commun.* **12**, 3770 (2021).

Acknowledgements

We extend our gratitude to patients and their families for partaking in this clinical trial and for their donations of tissue for research. We acknowledge the contributions of all the clinical research staff of the Hepatopancreatobiliary Service, the Precision Pathology Biobanking Center, and the Center for Molecular Oncology at MSK. This work was supported by a Stand Up to Cancer Convergence Award (to B.D.G. and V.P.B.), the Lustgarten Foundation (to B.D.G., and V.P.B.), the NIH U01 CA224175 Pancreatic Cancer Microenvironment Network Cancer Moonshot Award (to V.P.B.), the Damon Runyon Clinical Investigator Award (to V.P.B.), the Ben and Rose Cole Charitable PRIA Foundation (to V.P.B), the Mark Foundation ASPIRE Award (to B.D.G.) and the Pershing Square Sohn Prize-Mark Foundation Fellowship (to B.D.G.). M.Ł. is a Pew Biomedical Scholar. Services of the Integrated Genomics Core were funded by the NCI Cancer Center Support Grant (CCSG, P30 CA08748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology. The team at MSK sponsored the clinical trial. The imCORE Network, Genentech, BioNTech, Stand Up To Cancer, the Lustgarten Foundation, and the

National Cancer Institute Pancreatic Cancer Microenvironment Network funded the clinical trial and the biomarker studies.

Author information

Author notes

1. These authors contributed equally: Luis A. Rojas, Zachary Sethna

Authors and Affiliations

1. Immuno-Oncology Service, Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA

   Luis A. Rojas, Zachary Sethna, Pablo Guasp, Rebecca Yu, Adrienne Kaya Chandra, Theresa Waters, Jennifer Ruan, Masataka Amisaki, Abderezak Zebboudj, Zagaa Odgerel, George Payne & Vinod P. Balachandran

2. Hepatopancreatobiliary Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA

   Luis A. Rojas, Zachary Sethna, Kevin C. Soares, Cristina Olcese, Nan Pang, Erin Patterson, Pablo Guasp, Rebecca Yu, Adrienne Kaya Chandra, Theresa Waters, Jennifer Ruan, Masataka Amisaki, Abderezak Zebboudj, Zagaa Odgerel, George Payne, Alice C. Wei, Michael I. D'Angelica, T. Peter Kingham, William R. Jarnagin, Jeffrey Drebin & Vinod P. Balachandran

3. David M. Rubenstein Center for Pancreatic Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA

   Kevin C. Soares, Wungki Park, Alice C. Wei, Michael I. D'Angelica, T. Peter Kingham, William R. Jarnagin, Jeffrey Drebin, Eileen M. O'Reilly & Vinod P. Balachandran

4. Computational Oncology Service, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Jayon Lihm, Nicholas Ceglia, Alexander Chu & Benjamin D. Greenbaum

5. BioNTech, Mainz, Germany

Evelyna Derhovanessian, Felicitas Müller, Ugur Sahin & Özlem Türeci

6. Genentech, San Francisco, CA, USA

Ina Rhee, Mahesh Yadav & Ira Mellman

7. Center for Cell Engineering, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Anton Dobrin & Michel Sadelain

8. Immunology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Anton Dobrin & Michel Sadelain

9. Department of Oncological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Marta Łuksza

10. Department of Surgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA

Noah Cohen

11. Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Laura Tang & Olca Basturk

12. Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Mithat Gönen

13. Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Seth Katz & Richard Kinh Do

14. Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Andrew S. Epstein, Parisa Momtaz, Wungki Park, Ryan Sugarman, Anna M. Varghese, Elizabeth Won, Avni Desai & Eileen M. O'Reilly

15. Meyer Cancer Center, Weill Cornell Medicine, Weill Cornell Medical College, New York, NY, USA

Taha Merghoub & Jedd D. Wolchok

16. HI-TRON, Helmholtz Institute for Translational Oncology, Mainz, Germany

Özlem Türeci

17. Physiology, Biophysics and Systems Biology, Weill Cornell Medicine, Weill Cornell Medical College, New York, NY, USA

Benjamin D. Greenbaum

Contributions

V.P.B., J.D.W., T.M., Ö.T., I.M. and U.S. conceived the clinical trial concept. N.Cohen, E.O.R. and V.P.B. designed the clinical trial. V.P.B. and E.O.R. oversaw all aspects of clinical trial conduct as primary and co-primary investigators, respectively. A.S.E., P.M., W.P., R.S., A.M.V., E.W., A. Desai, E.O.R., K.C.S., A.C.W., M.I.D., T.P.K., W.R.J., J.D. and V.P.B. enrolled

patients to the clinical trial. K.C.S., A.C.W., M.I.D., T.P.K., W.R.J., J.D. and V.P.B. performed the surgeries. A.S.E., P.M., W.P., R.S., A.M.V., E.W., A. Desai, E.O.R. and V.P.B. administered the experimental medicines and evaluated toxicity. C.O. and N.P. led the MSK clinical trial research study team. S.K. and R.K.D performed the radiography assessments. O.B. and L.T. performed all the pathology assessments. L.A.R. performed the immunological assays with assistance from A. Dobrin, R.Y., A.K.C., T.W., J.R., M.A., P.G., A.Z., Z.O. and G.P. M.S. assisted with TCR cloning. Immunological analysis by IFNγ ELISpot was performed by E.D. (BioNTech). L.A.R., Z.S., B.D.G. and V.P.B. developed CloneTrack. Z.S. performed all computational analyses using CloneTrack. M.Ł., Z.S., L.A.R., B.D.G. and V.P.B. conceived and developed the neoantigen quality model. Z.S., J.L., A.C. and M.Ł. performed the clonality reconstructions. Z.S. performed the neoantigen quality assessments. Exome sequencing, RNA sequencing, identification of neoantigens, neoantigen selection, individualized mRNA vaccine design and manufacturing were performed by the BioNTech team coordinated by F.M. Z.S. and N.Ceglia performed the single-cell analyses. M.G. provided statistical oversight. The MSK team collected, analysed and interpreted all data, with critical feedback from all authors. L.A.R., Z.S., E.P., E.D., F.M., I.R., M.Y., T.M., J.D.W., I.M., U.S., Ö.T., B.D.G. and V.P.B. participated in critical discussions and manuscript writing. The funders had no role in the conduct of the trial or end point analyses. The authors assume responsibility for the completeness and accuracy of the data and for the fidelity of the trial to the protocol and statistical analysis plan.

Corresponding authors

Correspondence to Benjamin D. Greenbaum or Vinod P. Balachandran.

Ethics declarations

Competing interests

L.A.R. is an inventor of a patent related to oncolytic viral therapy (US20170051022A1). L.A.R., Z.S., B.D.G. and V.P.B. are inventors on patent applications related to work on antigen cross-reactivity

(PCT/US2023/011643) and tracking vaccine-expanded T cell clones. M.Ł., B.D.G. and V.P.B. are inventors on a patent application on neoantigen quality modelling (63/303,500). B.D.G. has received honoraria for speaking engagements from Merck, Bristol Meyers Squibb and Chugai Pharmaceuticals; has received research funding from Bristol Meyers Squibb, Merck, and ROME Therapeutics; and has been a compensated consultant for Darwin Health, Merck, PMV Pharma, Shennon Biotechnologies and Rome Therapeutics of which he is a co-founder. R.K.D. received one-time consulting fees from GE Healthcare and Bayer Healthcare. A.S.E. received royalties from Up-To-Date. A.M.V. discloses the following relationships (provision of services, uncompensated): Bristol-Myers Squibb, GlaxoSmithKline, Lilly Oncology, OBI Pharma, and Silenseed. M.S. has collaborative research agreements with Atara Biotherapeutics, Fate Therapeutics, Mnemo Therapeutics, and Takeda Pharmaceuticals. A.C.W. is a consultant for Histosonics and Biosapien; received honorarium from Medtronic and AstraZeneca; and received travel support from Intuitive Surgical. T.P.K. received a one-time consulting fee from Olympus Surgical. J.D. owns stock in Alnylam Pharmaceuticals, Arrowroot Acquisition, and Ionis Pharmaceuticals. W.P. receives grant and research support from Astellas, Merck, NIH/NCI, Parker Institute for Cancer Immunotherapy, and Break Through Cancer. E.M.O. receives grant and research support from Genentech/Roche, Celgene/Bristol-Myers Squibb, BioNTech, AstraZeneca, Arcus, Elicio, Parker Institute, and the NIH/NCI. T.M. is a co-founder and holds equity in IMVAQ Therapeutics; he is a consultant for Immunos Therapeutics, ImmunoGenesis and Pfizer; he has research support from Bristol-Myers Squibb, Surface Oncology, Kyn Therapeutics, Infinity Pharmaceuticals, Peregrine Pharmaceuticals, Adaptive Biotechnologies, Leap Therapeutics and Aprea; he has patents on applications related to work on oncolytic viral therapy, alphavirus-based vaccine, neoantigen modelling, CD40, GITR, OX40, PD-1 and CTLA-4. J.D.W. is a consultant for Apricity, CellCarta, Ascentage Pharma; AstraZeneca, Bicara Therapeutics, Boehringer Ingelheim, Bristol-Myers Squibb, Daiichi Sankyo, Dragonfly, Georgiamune, Imvaq, Larkspur, Psioxus, Recepta, Tizona, and Sellas. J.D.W. receives grant and research support from Bristol-Myers Squibb and Sephora. J.D.W. has equity in Apricity, Arsenal IO, Ascentage, Imvaq, Linneaus, Georgiamune, Maverick, and Tizona Therapeutics. V.P.B. has received honoraria for speaking engagements from Genentech, and research

support from Bristol-Myers Squibb and Genentech. Ö.T. and U.S. are co-founders, management board members and employees at BioNTech. E.D. and F.M. are employees at BioNTech. I.R., M.Y. and I.M. are employees at Genentech. The other authors (K.C.S., C.O., N.P., E.P., J.L., N. Ceglia, P.G., A.C., R.Y., A.K.C., T.W., J.R., M.A., A.Z., Z.O., G.P., A. Dobrin, N. Cohen, L.T., O.B., M.G., S.K., P.M., E.W., R.S., A. Desai, M.I.D. and W.R.J.) declare no competing interests.

Peer review

Peer review information

Nature thanks Patrick Ott, Robert Vonderheide and Neeha Zaidi for their contribution to the peer review of this work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Patient demographics, clinical characteristics, and toxicity.

(**A, B**) Demographics and clinical characteristics of all evaluable patients (n = 19) (**A**) and biomarker-evaluable patients (n = 16) stratified by autogene cevumeran responders and non-responders (**B**). (**C**) Frequency of grade 1 and 2 adverse events attributable to atezolizumab (**left**) and vaccine (**right**) in evaluable patients who received each drug. n = individual patients. Data are n (%) unless noted. P values by two-tailed Mann-Whitney test for numerical variables and Fisher's exact test for categorical variables.

Source Data

Extended Data Fig. 2 Immunogenic autogene cevumeran neoantigens.

(**A**) Frequency of immunogenic autogene cevumeran (vaccine) neoantigens in all patients (**left**) and in vaccine responders (**right**). (**B**) Number of immunogenic vaccine neoantigens per patient. (**C**) Number of non-synonymous mutations (**left**) and vaccine neoantigens (**right**) in vaccine responders and non-responders. (**D**) Vaccine-expanded T cell clones by CloneTrack in non-responders. *n* = neoantigens or patients as noted. *P* values by two-tailed Mann-Whitney test (**C**) and modified two-tailed Fisher's exact test (**D**).

Source Data

Extended Data Fig. 3 Autogene cevumeran–expanded T cell clones do not overlap with atezolizumab-expanded T cell clones.

(**A**) Cumulative percentage of atezolizumab-expanded T cells by CloneTrack in autogene cevumeran (vaccine) responders and non-responders. * = altered treatment schedules for patients 3, 4, 14 and 18. (**B**) Atezolizumab-expanded T cell clones by CloneTrack in vaccine responders and non-responders. Red box = atezolizumab responder. Blue line = trajectory of an individual atezolizumab-expanded clone. Black line = geometric mean trajectory of all atezolizumab-expanded clones; error bars = geometric standard deviation. Red line = cumulative percentage of all atezolizumab-expanded clones. Black rectangle/triangle = time of surgery; blue rectangle/triangle = time of atezolizumab; green rectangle/triangle = times of autogene cevumeran (vaccine); yellow bars = mFOLFIRINOX cycles. (**C**) Venn diagrams show overlap of vaccine-expanded clones (as identified in Fig. 2a) with atezolizumab-expanded clones in vaccine responders and non-responders. *n* = number of clones or individual patients as noted. *P* values by modified two-tailed Fisher's exact test (**B**).

Source Data

Extended Data Fig. 4 Autogene cevumeran–expanded T cell clones contain immunodominant neoantigen-specific T cells.

(**A, B**) Autogene cevumeran (vaccine) induced IFNγ production by ex vivo IFNγ ELISpot (assay schematic, Fig. 1e) and clonal expansion by CloneTrack (assay schematic, **B, top**). Briefly, for ELISpot, we analyzed each patient's PBMCs for specific T cells against all individual vaccine neoantigens post-atezolizumab/pre-vaccine, 1-3 weeks post-vaccine priming doses, and 5-6 weeks post-mFOLFIRINOX. Each patient's PBMCs were stimulated overnight with separate pools of overlapping peptides (15 amino acids long), each pool representing one of up to 20 neoantigens in vaccines, or with anti-CD3 antibody as a positive control, followed by measurement of IFNγ production by ELISpot. PBMCs incubated with media alone were used as a negative control. To track T cell clones, we identified vaccine-expanded T cell clones with CloneTrack. To identify if expanded clones contained immunodominant neoantigen-specific clones, in 4 of 8 responders, we stimulated vaccine-expanded PBMCs collected 1-3 weeks post-vaccine priming doses in vitro with computationally predicted minimal neopeptide pools (8-14 amino acids long) from 6 neoantigens that generated the highest per-patient magnitude response by ex vivo IFNγ ELISpot. We then purified $CD8^+$ T cells that either expressed or did not express the degranulation marker CD107a, identified clones with greater proportion of $CD107a^+$ versus $CD107a^-$ cells as in vitro neoepitope-activated clones, and examined overlap of in vitro neoepitope-activated to in vivo vaccine-expanded clones (Venn diagrams). For select patients, we further validated neoepitope-specificity by TCR cloning. (**A**) Flow cytometry gating strategies. (**B**) Ex vivo IFNγ ELISpot and T cell clonal expansion by CloneTrack in *n* = 3 of 4 patients tested (fourth patient in Fig. 2d). (**B, left**) Normalized ex vivo IFNγ ELISpot spot count per 1 x 10^6 PBMCs for each immunogenic neoantigen in patients 10, 11 and 5. In patient 5, one neoantigen induced a high magnitude T cell response, while patients 10 and 11 had polytopic high magnitude responses against 8 and 3 vaccine neoantigens respectively. Black lines = individual neoantigens; coloured lines = neoantigen pools. (**B, middle**): T cell clonal expansion in patients 10, 11 and 5 by CloneTrack. Green line = trajectory of an individual expanded clone. Black line = geometric mean trajectory of all expanded clones; error bars = geometric standard deviation. Red line = cumulative percentage of all expanded clones. Black triangle = time of surgery; blue triangle = time of atezolizumab; green triangles = times of autogene cevumeran (vaccine); yellow bars = mFOLFIRINOX cycles. Dotted black line = detection threshold. (**B, right**): In vitro neoantigen-

specific activation. Flow cytometry = CD107a expression on $CD8^+$ T cells stimulated with neopeptides or control (DMSO). Dot plots = number of $CD107a^+$ versus $CD107a^-$ cells per T cell clone. Each circle = individual T cell clone. Green/blue squares = clones also detected by CloneTrack. Diagonal: $CD107a^+$ clone frequency = $CD107a^-$ clone frequency. Venn diagram = vaccine-expanded and in vitro neoantigen-specific clonal overlap. For patient 11, immunodominant neoantigen-specific clones resided in a lower magnitude vaccine-expanded clonal pool. (**B, right, bottom**) TCR cloning, patient 10: 4-1BB expression on putative neoantigen-specific TCR-transduced $CD8^+$ T cells cultured with HLA-matched, neopeptide-pulsed antigen presenting cells (HLA-transduced K562 cells). *n* = number of clones. *P* values by modified two-tailed Fisher's exact test (**B, middle**), and by one-tailed binomial test with Bonferroni correction (**B, right**).

Source Data

Extended Data Fig. 5 Autogene cevumeran activates neoantigen-specific polyfunctional effector $CD8^+$ T cells.

(**A**) Uniform manifold approximation and projection (UMAP) plots of single peripheral blood T cells by single-cell RNA/TCR sequencing in *n* = 4 patients (patients 1, 10, 11 and 29) stratified by lineage (CD8 vs. CD4, **left**), patient (**middle**), and vaccine-expanded clones (**right**; expanded clones identified in Fig. 2a). T cells purified post-vaccine priming doses at time points indicated in Fig. 2b (inverted triangles). (**B, C**) UMAP plots of single peripheral blood $CD8^+$ T cells in patients 1, 10, 11 and 29 stratified by $CD8^+$ T cell naïve (*SELL, CCR7, IL7R, BCL2, PECAM1, TCF7, BACH2, LEF1*), dysfunctional (*TIGIT, TOX, LAG3, ENTPD1, CXCL13, HAVCR2, GZMB*), memory (*EOMES, GZMK, CXCR3, TCF1, ID3, STAT4, CCR7, SELL*) and effector (*PRF1, GZMB, GNLY, IFNγ, EOMES, ZEB2, E2F7, TBX21, PDCD1, CXCR3, FAS*) transcriptional phenotypes (**B**) and select individual phenotype defining markers (**C**). (**D**) Cytokine (IFNγ, TNFα) production in $CD8^+$ and $CD4^+$ T cells after post-vaccine bulk PBMC ex vivo rechallenge with pools of overlapping long neopeptides. (**E**) Cytokine (IFNγ, TNFα) production and degranulation (CD107a) by $CD8^+$ T cells after bulk PBMC ex vivo rechallenge with predicted minimal neopeptides. Flow cytometry

gating strategies as in Extended Data Fig. 4a. (**F**) Peripheral blood CD8$^+$ T cell proliferation (Ki67) and activation (PD-1, LAG-3, TIM-3, HLA-DR) pre- and serially post-atezolizumab, vaccine and mFOLFIRINOX by flow cytometry in $n = 7$ of 8 responders and $n = 7$ of 8 non-responders with available data. n = number of patients.

Source Data

Extended Data Fig. 6 Post-mFOLFIRINOX autogene cevumeran booster re-expands primed T cell clones.

(**A**) CloneTrack-identified autogene cevumeran (vaccine) expanded T cell clones with vaccine priming, mFOLFIRINOX chemotherapy (chemo) and vaccine booster. Data shown for patients that had detectable vaccine-expanded clones and received the vaccine booster. (**B**) Percentage of vaccine-expanded T cell clones that are detectable and re-expand with vaccine booster in responders. n = number of patients or clones.

Source Data

Extended Data Fig. 7 Autogene cevumeran responders evidence lower post-vaccination serum CA19-9, equivalent chemotherapy doses and comparable intratumoural T cells.

(**A, B**) Fraction of biomarker-evaluable patients with detectable CA19-9 (**A**) or circulating tumour DNA (ctDNA) (**B**) in the peripheral blood at diagnosis. Longitudinal serum CA19-9 levels in autogene cevumeran (vaccine) responders and non-responders (**A, bottom**). Circle = mean, error bars = standard error of the mean (**C**) Recurrence-free survival (RFS) stratified by atezolizumab response (**left**) and median primary tumour size (**right**). (**D**) Number of cycles of mFOLFIRINOX (chemotherapy) in vaccine responders and non-responders. (**E**) Intratumoural T cell infiltration in autogene cevumeran responders and non-responders. n = individual patients. P values by two-tailed Mann Whitney test (**A**), and log-rank test (**C**).

Source Data

Extended Data Fig. 8 Autogene cevumeran responders and non-responders have equivalent humoral and cellular responses to an unrelated mRNA vaccine.

(**Top left**) Anti-SARS-CoV-2 spike IgG in sera of vaccine responders and non-responders before (pre), after priming (post-prime) and after booster (post-boost early and late) doses of COVID-19 mRNA vaccine (Pfizer-BioNTech BNT162b2 or Moderna Spikevax). (**Top right**) Serial PBMCs pre- and post-COVID-19 mRNA vaccination were stimulated with SARS CoV-2 peptide pools. SARS CoV-2-specific IFNγ and/or TNFα production by all T cells was measured by flow cytometry. Composite data (**top right**) with representative gating (**bottom**) are shown. Flow cytometry pre-gated on $CD3^{+}$ $CD56^{-}$ cells. Circle = mean, error bars = standard error of the mean; *n* = individual patients; data shown for all patients with available samples. *P* values by two-tailed Wilcoxon matched-pairs signed rank test.

Source Data

Extended Data Fig. 9 Autogene cevumeran responders and non-responders have equivalent frequencies of circulating immune cells.

(**A-C**) Longitudinal frequencies of regulatory T cells (**A**), innate (**B**) and adaptive (**C**) immune cells in the peripheral blood of autogene cevumeran responders and non-responders during vaccination. Analyses in $n = 13$ patients with identical study-specified treatment schedules and thus eligible for direct comparison. Circle = mean, error bars = standard error of the mean; *n* = individual patients. *P* values by two-tailed Mann Whitney test.

Source Data

Extended Data Fig. 10 Tumour clonality and neoantigen quality in biomarker-evaluable patients, and mutated *TP53* in a disappearing liver lesion.

(**A**) Shannon entropy (*S*) of tumour clones in autogene cevumeran (vaccine) responders and non-responders. (**B**) Receiver operating curve indicating the ability of neoantigen quality as a continuous variable to classify vaccine neoantigens as an inducer or non-inducer of an IFNγ ELISpot response. Dotted line = all neoantigens included in individualized mRNA vaccines. (**C**) Digital droplet PCR showing number of wild-type (WT) *TP53*, or R175H mutated *TP53* droplets in liver lesion from patient 29 (Fig. 4), with positive and negative (gDNA) controls. *n* = individual patients (**A**) or neoantigens (**B**). *P* values by two-tailed Mann-Whitney test.

Source Data

Extended Data Table 1 Primary Safety Endpoint
Extended Data Table 2 Repertoire rescaling for CloneTrack

Supplementary information

Supplementary Data 1

Clinical trial protocol.

Reporting Summary

Supplementary Tables 1–4

Supplementary Table 5

Source data

Source Data Fig. 1

Source Data Fig. 2

Source Data Fig. 4

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 2

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 4

Source Data Extended Data Fig. 5

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 9

Source Data Extended Data Fig. 10

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Rojas, L.A., Sethna, Z., Soares, K.C. *et al.* Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. *Nature* **618**, 144–150 (2023). https://doi.org/10.1038/s41586-023-06063-y

- Received: 10 January 2023
- Accepted: 06 April 2023
- Published: 10 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06063-y

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

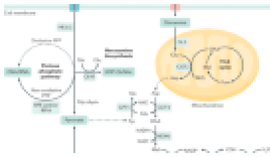
Provided by the Springer Nature SharedIt content-sharing initiative

Vaccine boosts T cells that target pancreatic tumours

- Amanda L. Huff
- Neeha Zaidi

Nature News & Views 10 May 2023

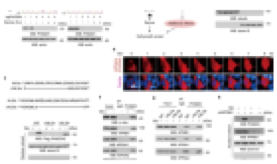
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06063-y



Harnessing metabolic dependencies in pancreatic cancers

19 March 2021

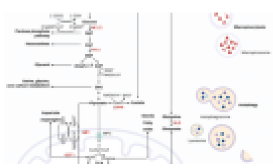
Joel Encarnación-Rosado & Alec C. Kimmelman



The alternative activity of nuclear PHGDH contributes to tumour growth under nutrient stress

18 October 2021

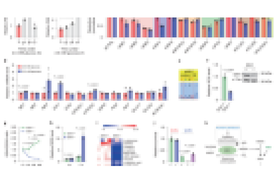
Chunmin Ma, Ke Zheng, … Yuhui Jiang



Interplays of glucose metabolism and KRAS mutation in pancreatic ductal adenocarcinoma

24 September 2022

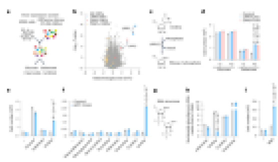
Yu-Huei Liu, Chun-Mei Hu, … Wen-Hwa Lee



[Limited nutrient availability in the tumor microenvironment renders pancreatic tumors sensitive to allosteric IDH1 inhibitors]

09 June 2022

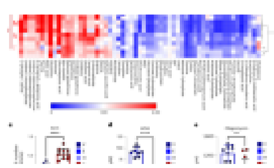
Ali Vaziri-Gohar, Joel Cassel, … Jordan M. Winter



[Salvage of ribose from uridine or RNA supports glycolysis in nutrient-limited conditions]

17 May 2023

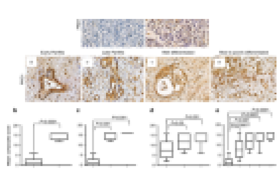
Owen S. Skinner, Joan Blanco-Fernández, … Alexis A. Jourdain



[Differential integrated stress response and asparagine production drive symbiosis and therapy resistance of pancreatic adenocarcinoma cells]

21 November 2022

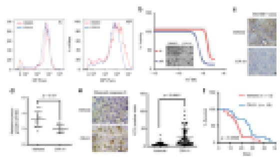
Christopher J. Halbrook, Galloway Thurston, … Costas A. Lyssiotis



[Protein kinase D1 regulates metabolic switch in pancreatic cancer via modulation of mTORC1]

10 December 2019

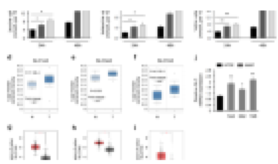
Sonam Kumari, Sheema Khan, … Meena Jaggi



[eIF4A supports an oncogenic translation program in pancreatic ductal adenocarcinoma]

13 November 2019

Karina Chan, Francis Robert, … Iok In Christine Chio



[Branched-chain amino acids sustain pancreatic cancer growth by regulating lipid metabolism]

29 November 2019

Ji Hyeon Lee, Young-ra Cho, … Jaekyoung Son

- Article
- [Open Access]
- [Published: 17 May 2023]

Uridine-derived ribose fuels glucose-restricted pancreatic cancer

- Zeribe C. Nwosu,
- Matthew H. Ward,
- Peter Sajjakulnukit,
- Pawan Poudel,
- Chanthirika Ragulan,
- Steven Kasperek,
- Megan Radyk,
- Damien Sutton,
- Rosa E. Menjivar,
- Anthony Andren,
- Juan J. Apiz-Saab,
- Zachary Tolstyka,
- Kristee Brown,
- Ho-Joon Lee,
- Lindsey N. Dzierozynski,
- Xi He,
- Hari PS,
- Julia Ugras,
- Gift Nyamundanda,
- Li Zhang,
- Christopher J. Halbrook,
- Eileen S. Carpenter,
- Jiaqi Shi,
- Leah P. Shriver,
- Gary J. Patti,
- Alexander Muir,
- Marina Pasca di Magliano,
- Anguraj Sadanandam &
- …
- Costas A. Lyssiotis

Show authors

Nature **volume 618**, pages 151–158 (2023)

- 15k Accesses
- 423 Altmetric
- Metrics details

Abstract

Pancreatic ductal adenocarcinoma (PDA) is a lethal disease notoriously resistant to therapy[1,2]. This is mediated in part by a complex tumour microenvironment[3], low vascularity[4], and metabolic aberrations[5,6]. Although altered metabolism drives tumour progression, the spectrum of metabolites used as nutrients by PDA remains largely unknown. Here we identified uridine as a fuel for PDA in glucose-deprived conditions by assessing how more than 175 metabolites impacted metabolic activity in 21 pancreatic cell lines under nutrient restriction. Uridine utilization strongly correlated with the expression of uridine phosphorylase 1 (UPP1), which we demonstrate liberates uridine-derived ribose to fuel central carbon metabolism and thereby support redox balance, survival and proliferation in glucose-restricted PDA cells. In PDA, *UPP1* is regulated by KRAS–MAPK signalling and is augmented by nutrient restriction. Consistently, tumours expressed high *UPP1* compared with non-tumoural tissues, and *UPP1* expression correlated with poor survival in cohorts of patients with PDA. Uridine is available in the tumour microenvironment, and we demonstrated that uridine-derived ribose is actively catabolized in tumours. Finally, UPP1 deletion restricted the ability of PDA cells to use uridine and blunted tumour growth in immunocompetent mouse models. Our data identify uridine utilization as an important compensatory metabolic process in nutrient-deprived PDA cells, suggesting a novel metabolic axis for PDA therapy.

Main

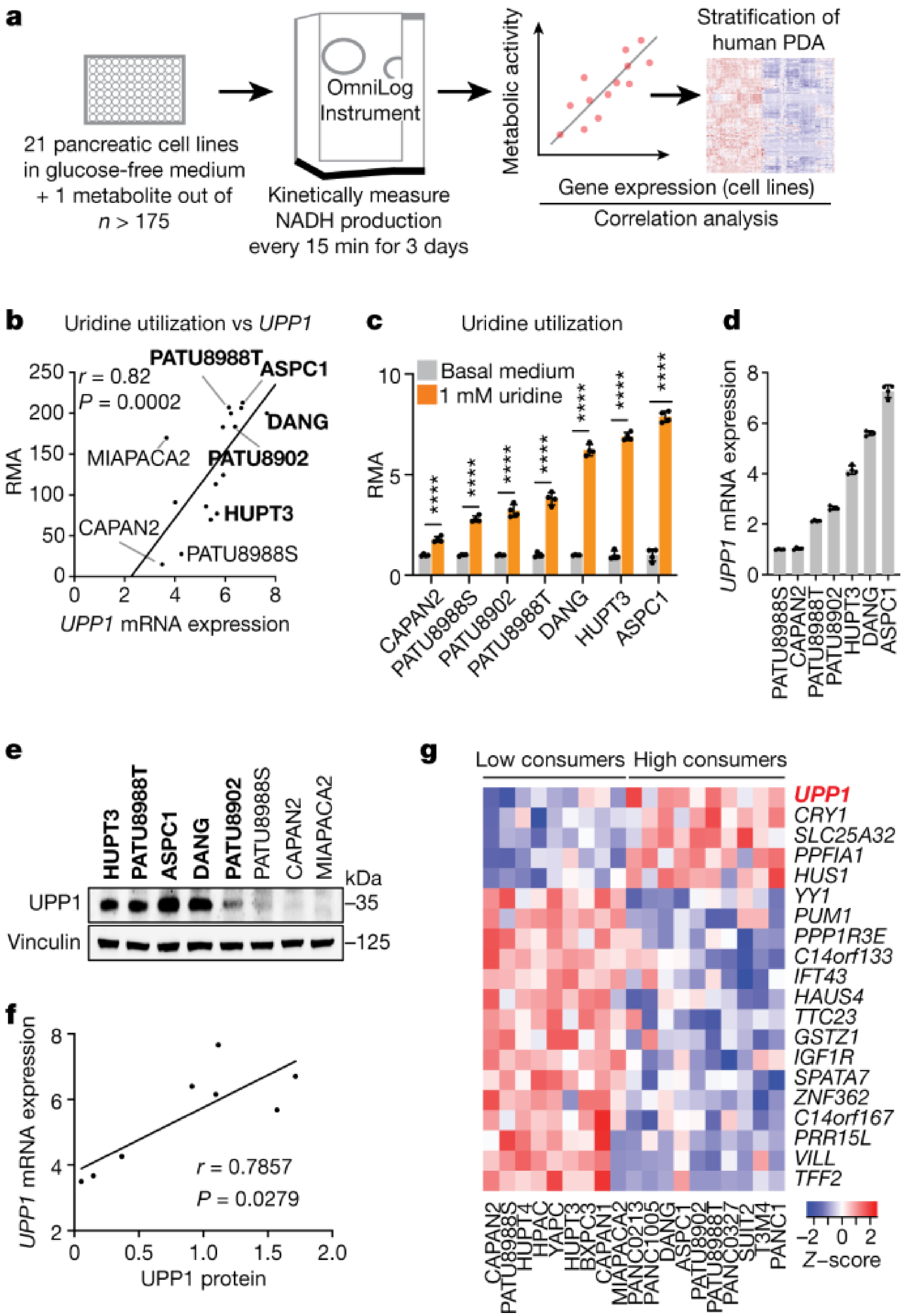
PDA remains one of the deadliest cancers[1,2]. The PDA tumour microenvironment (TME) is a major contributor to this lethality, and is characterized by abundant immune cell infiltration, expansion of stromal

fibroblasts and the associated deposition of extracellular matrix. This leads to an increase in interstitial fluid pressure and the collapse of arterioles and capillaries[3,4,7]. These phenomena collectively contribute to low oxygen saturation, therapeutic resistance, metabolic alterations, and heterogeneity within the tumour at the cellular level[5,8,9]. PDA cells surviving in such nutrient and oxygen deregulated TME exhibit metabolic adaptations that increase their scavenging and catabolic capabilities[10,11,12,13]. In addition, recent studies have defined tumour-extrinsic nutrient sources for PDA, including extracellular matrix, immune, and stromal-derived metabolites[14,15,16]. While these studies uncovered discrete nutrient inputs, comprehensive screens with the power to identify many such nutrient drivers and mechanisms have not been performed previously.

Nutrient-deprived PDA consumes uridine

To screen for metabolites that fuel metabolism in nutrient-deprived PDA cells, we applied the Biolog phenotypic screening platform on 19 human PDA cell lines and 2 immortalized, non-malignant pancreas cell lines (human pancreatic stellate cells and human pancreatic nestin-expressing cells) (Fig. 1a). We used the screen to assess cellular ability to capture and metabolize more than 175 nutrients in a 96-well arrayed format under nutrient-limiting conditions (0 mM glucose, 0.3 mM glutamine and 5% dialysed fetal bovine serum (FBS)). The nutrient panel included carbon energy and nitrogen substrates (Supplementary Table 1). Metabolic activity was assessed by monitoring the reduction of a tetrazolium-based dye, a readout of cellular reducing potential, every 15 min for approximately 3 days (Fig. 1a and Extended Data Fig. 1a). Analyses of nutrient consumption profiles revealed several metabolites that, in the absence of glucose, were utilized at similar levels as the glucose positive control (Extended Data Fig. 1b). For example, adenosine, uridine and several sugars were utilized by most of the cell lines.

Fig. 1: Profiling of metabolite utilization in PDA cells identifies uridine.



a
21 pancreatic cell lines in glucose-free medium + 1 metabolite out of n > 175
OmniLog Instrument
Kinetically measure NADH production every 15 min for 3 days
Metabolic activity
Gene expression (cell lines)
Correlation analysis
Stratification of human PDA
b
Uridine utilization vs UPP1
r = 0.82
P = 0.0002
RMA
UPP1 mRNA expression
PATU8988T
ASPC1
DANG
MIAPACA2
PATU8902
HUPT3
CAPAN2
PATU8988S
c
Uridine utilization
Basal medium
1 mM uridine
RMA
CAPAN2
PATU8988S
PATU8902
PATU8988T
DANG
HUPT3
ASPC1
d
UPP1 mRNA expression
PATU8988S
CAPAN2
PATU8988T
PATU8902
HUPT3
DANG
ASPC1
e
HUPT3
PATU8988T
ASPC1
DANG
PATU8902
PATU8988S
CAPAN2
MIAPACA2
kDa
UPP1
–35
Vinculin
–125
f
UPP1 mRNA expression
UPP1 protein
r = 0.7857
P = 0.0279
g
Low consumers
High consumers
UPP1
CRY1
SLC25A32
PPFIA1
HUS1
YY1
PUM1
PPP1R3E
C14orf133
IFT43
HAUS4
TTC23
GSTZ1
IGF1R
SPATA7
ZNF362
C14orf167
PRR15L
VILL
TFF2
CAPAN2
PATU8988S
HUPT4
HPAC
YAPC
HUPT3
BXPC3
CAPAN1
MIAPACA2
PANC0213
PANC1005
DANG
ASPC1
PATU8902
PATU8988T
PANC0327
SUIT2
T3M4
PANC1
Z-score

a, Scheme of the nutrient metabolism screening assay and the correlation with gene expression in PDA cell lines and tumours. **b**, Spearman correlation (*r*) between the normalized relative metabolic activity (RMA) for uridine catabolism in the screening data and *UPP1* mRNA expression data from an independent dataset[17] (16 PDA cell lines). *UPP1*-high cell lines are shown in bold. **c**, The RMA in a subset of PDA cell lines following supplementation with 1 mM uridine for 3 days in glucose-free condition. **d**, Quantitative PCR (qPCR) validation of *UPP1* mRNA expression in a subset of PDA cell lines. **e**, Immunoblot showing basal UPP1 expression in PDA cell lines. Blots are representative of three technical replicates with similar results. **f**, Spearman correlation (*r*) between protein densitometry analysis of the blot in **e** and *UPP1* mRNA expression in the eight PDA cell lines highlighted in **e**. **g**, Top 20 genes differentially expressed by the PDA cell lines that were identified as uridine-high consumers/metabolizers compared with uridine-low consumers from the nutrient metabolism screen. Data source: Cancer Cell Line Encyclopedia (CCLE). Data in **c**,**d** are mean ± s.d. See Methods ‘Statistics and reproducibility’ section for additional information.

Uridine consumption correlates with UPP1

To select lead metabolites for investigation, we correlated metabolite utilization patterns to the expression of metabolism-associated genes using a public dataset[17,18]. From the top metabolite and gene correlation pairs (Extended Data Fig. 1c), we pursued uridine and uridine phosphorylase 1 (UPP1) ($r = 0.82$, $P = 0.0002$; Fig. 1b) for the following reasons. First, *UPP1* expression correlated positively with the metabolic activity from its known substrate, uridine. Second, our in vitro validation showed that all tested PDA cells utilized uridine, although to varying degrees (Fig. 1c and Extended Data Fig. 1b), suggesting that it is a broadly used metabolic fuel. Third, contrary to our expectation that nutrients used in the absence of glucose would be carbohydrates, uridine was unusual in that it is a nucleoside. Finally, to our knowledge, UPP1-mediated uridine metabolism is unexplored in the context of PDA.

We further confirmed the correlation of uridine catabolism and UPP1 expression by mRNA and protein analyses (Fig. 1d–f). To determine the

specificity of this association, we assessed the correlation of *UPP1* expression to other nucleosides in the Biolog screen. Although both inosine and adenosine were readily catabolized, their utilization was not correlated with *UPP1* expression. Thymidine was neither actively metabolized nor correlated, when compared with negative controls (Extended Data Fig. 1d). These results indicate that the association between *UPP1* expression and uridine catabolism is robust and specific.

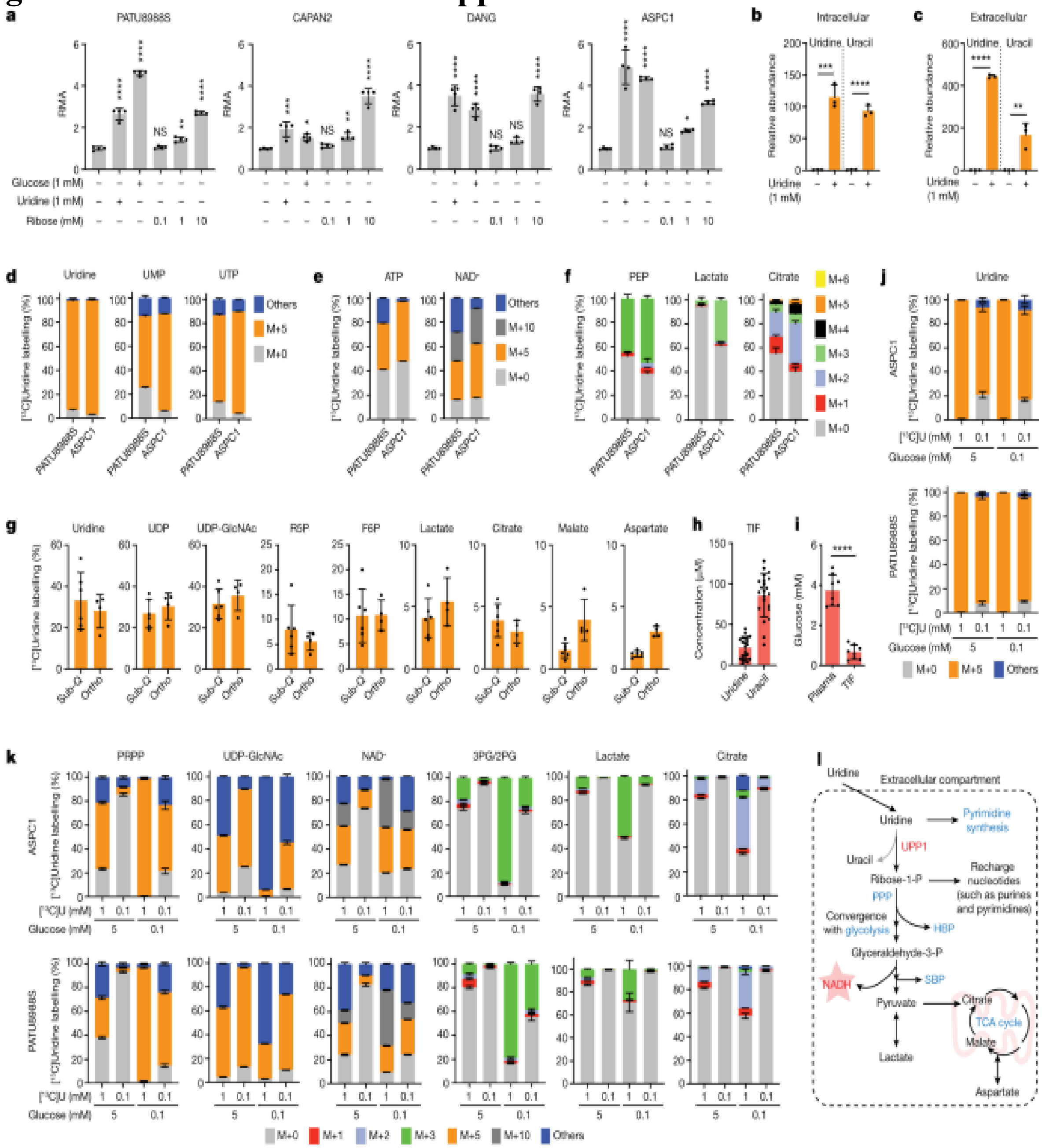
Next, we rank-sorted the PDA cell lines into high and low uridine metabolizers on the basis of the Biolog data, and identified more than 700 differentially expressed genes ($P < 0.05$) between these groups using the Cancer Cell Line Encyclopedia (CCLE) data[19] (Supplementary Table 2). Consistent with our previous correlation analysis in a different dataset (Extended Data Fig. 1c), UPP1 was the top gene in uridine-high consumers/metabolizers in the CCLE dataset (Fig. 1g). Pathway analysis of the upregulated genes in the uridine-high metabolizers (that is, *UPP1*-high cell lines) showed upregulation of endocytosis and several inflammation and/or immune-related pathways, notably NFκB signalling (Supplementary Table 3). We also observed that *UPP1*-high tumours exhibit higher expression of glycolysis genes (Extended Data Fig. 1e), indicating a potential link between the UPP1–uridine axis and energy metabolism. By contrast, *UPP1*-high cell lines and *UPP1*-high PDA tumours from patients displayed a profound downregulation of other metabolic pathways (Extended Data Fig. 1f,g), notably amino acid, fatty acid and glutathione metabolism.

Uridine-derived ribose fuels metabolism

Our screen (Fig. 1a) was performed in glucose-free medium to reveal both metabolites whose use would otherwise be overshadowed by glucose, and the carbon sources that could act in place of glucose. Thus, we next directly assessed the metabolic activity of equimolar glucose and uridine. Across four PDA cell lines, uridine and glucose fuelled metabolism to a similar degree (Fig. 2a). Previous reports have also documented that uridine can substitute for glucose by supporting nucleotide metabolism[20,21,22,23,24,25,26]. However, our screening data illustrate that uridine supplementation increases the cellular reducing potential. Together with the observed connection with

UPP1, which catalyses the cleavage of uridine to ribose-1-phosphate and uracil, we hypothesized that the UPP1-liberated ribose is recycled into central carbon metabolism to support cellular reducing potential. To test this hypothesis, we supplemented glucose-deprived cells with ribose, a cell-permeable substitute for ribose 1-phosphate. Indeed, similar to exogenous uridine, ribose supplementation fuelled the reducing potential (Fig. 2a).

Fig. 2: Uridine-derived ribose supports nutrient-restricted PDA.



a, RMA of four PDA cell lines supplemented with glucose, uridine or ribose under nutrient-limiting culture conditions (0 mM glucose, 0.3 mM glutamine and 5% dialysed FBS). **b**,**c**, Intracellular and extracellular uridine and uracil after 24 h culture of PATU8988S cells in medium with no glucose and 10% dialysed FBS with or without 1 mM uridine as measured by LC–MS. **d**–**f**, Mass isotopologue distribution of carbon derived from [$^{13}C_5$]uridine in uridine, UMP and UTP (**d**), ATP and NAD^+ (**e**) and phosphoenolpyruvate (PEP), lactate and citrate (**f**) in the indicated cell lines after 24 h culture with 1 mM uridine. **g**, Isotope tracing showing [$^{13}C_5$]uridine-derived carbon labelling in subcutaneous (sub-Q) or orthotopically (ortho) implanted KPC 7940b tumours collected 1 h after injecting the mice with 0.2 M [$^{13}C_5$]uridine. F6P, fructose-6-phosphate; R5P, ribose-5-phosphate; UDP-GlcNAc, uridine diphosphate *N*-acetylglucosamine; **h**, Absolute quantification via metabolomics of uridine and uracil concentration in the TIF of orthotopic PDA tumours from syngeneic mouse KPC cells. **i**, Absolute quantification via metabolomics of glucose concentration in the pancreatic TIF and plasma of mice orthotopically implanted with KPC 7490b syngeneic tumours. **j**,**k**, Mass isotopologue distribution of [$^{13}C_5$]uridine ribose-derived carbon after 24 h culture of ASPC1 and PATU8988S cells in medium supplemented with 1 mM or 0.1 mM uridine and each with 5 mM or 0.1 mM glucose. **j**, Uridine. **k**, PRPP, UDP-GlcNAc, NAD^+, 3-phosphoglycerate/2-phosphoglycerate (3PG/2PG), lactate and citrate. PRPP, phosphoribosyl pyrophosphate. **l**. Schematic depicting the fate of uridine-derived ribose carbon in PDA cells actively catabolizing uridine. Glyceraldehyde-3-P, glyceraldehyde-3-phosphate; HBP, hexosamine biosynthetic pathway; PPP, pentose phosphate pathway; ribose-1-P, ribose-1-phosphate; SBP, serine biosynthesis pathway. See Methods 'Statistics and reproducibility' section for additional information.

In our initial screen, glutamine concentration was also intentionally low (0.3 mM), as it is an important anaplerotic substrate that fuels tricarboxylic acid (TCA) cycle and proliferation in PDA[27]. Uridine potentiated reducing potential with or without glutamine and had a greater effect when glutamine was present (Extended Data Fig. 2a). Together, these data suggest that uridine and glucose similarly fuel central carbon metabolism distinctly from glutamine.

Next, we provided uridine to the UPP1-low PATU8988S and the UPP1-high DANG cell lines under glucose deprivation and applied liquid chromatography–mass spectrometry (LC–MS)-based metabolomics[28]. In both cell lines, uridine supplementation led to increased levels of glycolytic intermediates and lactate secretion (suggesting glycolytic flux), uridine derivatives (suggesting overflow metabolism), amino acids (indicative of increased anabolism) and TCA cycle intermediates (suggesting more mitochondrial activity) (Extended Data Fig. 2b–e). Moreover, supplementation with uridine led to a marked accumulation of intracellular uridine and over 100-fold increase in uracil content in the medium (Fig. 2b,c and Extended Data Fig. 2f,g), consistent with uridine capture, ribose extraction and uracil release. Notably, the intracellular uracil concentration increased by a similar amount of uridine, reflective of direct conversion of substrate to product (Fig. 2b and Extended Data Fig. 2f). Collectively, these profiling efforts support a model in which uridine is catabolized to broadly fuel PDA cell metabolism.

To precisely delineate how uridine is metabolized, we used LC–MS to trace the metabolic fate of isotopically labelled uridine ([$^{13}C_5$]uridine with uniformly labelled ribose carbon)[29] in PATU8988S (UPP1-low) and ASPC1 (UPP1-high) cell lines. Both cell lines demonstrated high uridine, UMP and UTP labelling (over 90%), as indicated by M+5 from ribose (Fig. 2d). Other nucleotides, such as ATP, AMP and ADP (all M+5), as well as NAD^+ (M+5, M+10) were also labelled (Fig. 2e and Extended Data Fig. 2h), demonstrating the use of uridine-derived ribose for ribosylation of the adenine base. Also labelled were glycolytic (PEP, pyruvate and lactate), PPP (X5P and ribose-5-phosphate), hexosamine biosynthetic pathway (UDP-GlcNAc) and TCA cycle intermediates (malate and citrate), as well as non-essential amino acids (aspartate, glutamate and serine) and oxidized glutathione (Fig. 2f and Extended Data Fig. 2h).

To determine the relevance of uridine metabolism for pancreatic tumours in vivo, we implanted mouse syngeneic pancreatic cancer cells into the pancreas of immunocompetent hosts to establish tumours. Then, we injected these mice with [$^{13}C_5$]uridine, and collected tumour tissues after 1 h for LC–MS analysis. Indeed, there was a robust uptake of uridine by the tumours, with almost 30% of the uridine pool being labelled. We observed the M+5

label in pyrimidine and purine species (that is, ribose salvage) as well as in glycolytic and TCA cycle intermediates (Fig. 2g and Extended Data Fig. 3a). ^{13}C incorporation into the TCA cycle was low in vivo, presumably owing to the short duration of labelling. Nearly identical results to those from orthotopic studies were observed in subcutaneous tumours from immune-competent mice. These results confirm that PDA catabolizes uridine in vivo.

In parallel, we collected tumour interstitial fluid (TIF) from independent orthotopic allograft tumours and quantified bulk uridine and glucose concentrations by LC–MS. Uridine and glucose were present in the low and high micromolar concentration range, respectively (Fig. 2h,i), similar to previous findings[30]. To determine how physiological concentrations of glucose and uridine affected uridine metabolism, we grew two PDA lines in 5 mM or 0.1 mM glucose and 0.1 mM or 1 mM [$^{13}C_5$]uridine and analysed labelling patterns. First, the labelling of uridine was nearly 100% (Fig. 2j). At the low equimolar concentration (0.1 mM uridine and glucose), uridine carbon contributed to several metabolites in the PPP, PRPP (involved in nucleotide biosynthesis), NAD^+, glycolysis, and TCA cycle, exceeding 50% enrichment in some cases(Fig. 2k and Extended Data Fig. 3b). When glucose was 50-fold higher (5 mM), uridine carbon contributed to a much lower level to metabolite labelling, whereas at lower glucose concentrations, uridine carbon dominated, consistent with competition for these two carbon sources into the same pathways. Further isotope tracing using the exact TIF concentrations of uridine and glucose showed that both human (ASPC1) and mouse (MT3-2D) cells incorporate uridine into central carbon metabolism (Extended Data Fig. 3c,d). We confirmed these results in four human PDA lines with the tetrazolium assay: uridine supported bioenergetics at physiological (Extended Data Fig. 4a,b) but not at elevated glucose levels (Extended Data Fig. 4c).

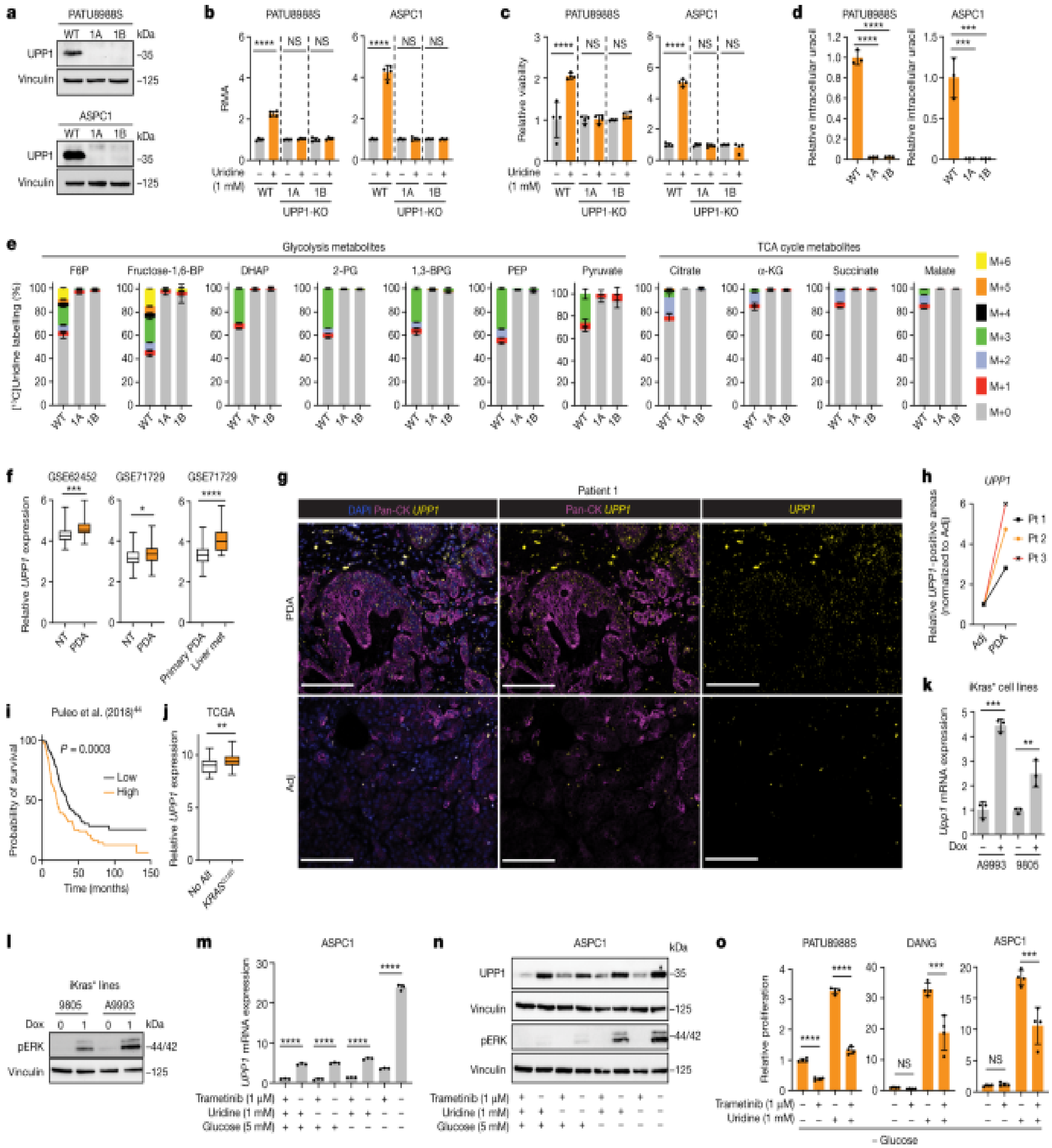
Our data suggest that uridine yields ribose-1-phosphate via UPP1 to fuel both catabolic and biosynthetic metabolism. Ribose-1-phosphate can be converted to the PPP product ribose-5-phosphate by phosphoglucomutase 2 (PGM2) to enter nucleotide biosynthesis. Alternately, cells can convert uridine to UMP via uridine-cytidine kinase (UCK1/2) in the pyrimidine salvage pathway (Extended Data Fig. 4d). We found that *PGM2* and *UCK2* are high in PDA and *UCK1* is low, but these genes were largely uncorrelated

with *UPP1* (Extended Data Fig. 4e–h). Being the most upregulated, we tested PGM2 by western blot and found it to be expressed in most PDA cells but uncorrelated with UPP1 (Extended Data Fig. 4i). Inhibition of the three genes using short interfering RNA (siRNA) showed that only *PGM2* knockdown suppressed the uridine-mediated rescue of metabolic activity following glucose deprivation (Extended Data Fig. 4j–l). Together, these data support our model in which uridine catabolism converges with central carbon metabolism, and they also reveal that exogenous uridine fuels PDA metabolism in a similar way to glucose, supplying carbon for redox, nucleotide, amino acid and glycosylation metabolite biosynthesis (Fig. 2l).

UPP1 provides uridine-derived ribose

To confirm the role of UPP1 in uridine catabolism, we knocked out UPP1 (UPP1-KO) using CRISPR–Cas9 in the PATU8988S (UPP1-low) and ASPC1 (UPP1-high) human PDA cell lines and validated two independent clones per cell line (Fig. 3a). In these knockout lines, the ability of uridine to rescue NADH production in the absence of glucose (Fig. 3b) or cellular bioenergetics, read out by ATP-based viability (Fig. 3c), was abolished. Consistent with the blockade of uridine catabolism, metabolomics showed that UPP1-KO cell lines displayed an increase in intracellular and extracellular uridine (Extended Data Fig. 5a,b) accompanied by a marked drop in intracellular and extracellular uracil (Fig. 3d and Extended Data Fig. 5b). Furthermore, UPP1-KO broadly altered the intracellular metabolome of both cell lines (Extended Data Fig. 5c). Notebly, ASPC1 cells are highly sensitive to glucose deprivation in combination with UPP1-KO, thus metabolomics in this cell was performed after 6 h when assessing the knockout effect.

Fig. 3: KRAS-driven UPP1 liberates ribose and is increased in PDA.



a, Western blot validation of UPP1-KO in human PDA cell lines. 1A and 1B denote UPP1-KO cells. WT, wild type. **b**,**c**, RMA from tetrazolium assay showing uridine-derived reducing potential (**b**) and CellTiter Glo showing ATP production (**c**) in UPP1-KO and wild-type PATU8988S and ASPC1 cells cultured with or without 1 mM uridine for 48 h. **d**, Relative intracellular uracil as determined by LC–MS in wild-type and UPP1-KO human PDA clones cultured with 1 mM uridine for 24 h (PATU8988S) or 6 h (ASPC1). **e**, Mass isotopologue distribution of 1 mM [$^{13}C_5$]uridine-derived

carbon in glycolysis and TCA cycle metabolites in wild-type or UPP1-KO ASPC1 cells after 6 h. α-KG, α-ketoglutarate; 1,3-BPG, 1,3-bisphosphoglycerate; DHAP, dihydroxyacetone phosphate; fructose-1,6-BP, fructose-1,6-bisphosphate. **f**, *UPP1* mRNA expression in PDA tumours and non-tumoural pancreas tissues in microarray datasets. Liver met, liver metastasis; NT, non-tumour tissue. **g**,**h**, RNAscope showing representative *UPP1* mRNA expression in tumour and adjacent normal tissue (adj) sections (**g**) and quantification from three patients (Pt 1–3) with PDA (**h**). Scale bars, 100 μm. **i**, Kaplan–Meier overall survival analysis (log-rank test) based on ranked *UPP1* expression in the PDA dataset published previously[44]. **j**, Comparison of *UPP1* mRNA expression in human PDA tumours annotated as $KRAS^{G12D}$ or with no alteration (No Alt) in *KRAS* from the TCGA dataset. **k**, qPCR data showing *Upp1* expression in mouse cell lines (A9993 and 9805) with doxycycline-inducible oncogenic *Kras* (iKras*). **l**, Western blot validation of MAPK pathway induction as indicated by phosphorylated ERK (pERK) in the iKras* cell lines. **m**,**n**, qPCR for *UPP1* mRNA (**m**) and western blot for pERK and UPP1 (**n**) in ASPC1 cells cultured with or without 5 mM glucose, 1 mM uridine and 1 μM trametinib for 48 h. **o**, MTT assay showing relative proliferation of PDA cell lines with 1.25 μM trametinib and 1 mM uridine in the absence of glucose. See Methods 'Statistics and reproducibility' section for additional information.

We next used our isotope tracing metabolomics platform to determine the effect of UPP1-KO on uridine catabolism in PATU8988S and ASPC1 cells (Fig. 3e and Extended Data Fig. 5d–k). The ribose-labelled [$^{13}C_5$]uridine tracing showed similar fractional enrichment of the intracellular uridine pool in the UPP1-KO and control cells (Extended Data Fig. 5d,e), indicative of unchanged, steady-state uridine uptake. By contrast, and consistent with our model, flux of uridine ribose-derived carbon into glycolysis (Fig. 3e and Extended Data Fig. 5f), TCA cycle-associated metabolites (Fig. 3e and Extended Data Fig. 5g), non-essential amino acids, oxidized glutathione and UDP-GlcNAc (in glycosylation), was holistically blocked or suppressed (Extended Data Fig. 5h,i). Recycling of uridine-derived ribose was also completely blocked in the UPP1-KO cells, as evidenced by the absence of carbon labelling in NAD^+ and the bioenergetic metabolites AMP, ADP and ATP (Extended Data Fig. 5j,k). Together, these results reveal that the UPP1-mediated catabolism of uridine is indispensable for the utilization of uridine

to support reducing potential, bioenergetics and cell proliferation, and provide a detailed molecular confirmation that UPP1 directly controls the utilization of uridine-derived ribose in PDA cells.

High UPP1 in PDA predicts poor survival

To further assess the relevance of UPP1 in PDA tumours, we next analysed its expression in publicly available human PDA datasets. We found that *UPP1* is highly expressed in PDA tumours compared with non-tumoural samples, and also in liver metastasis compared with primary tumours (Fig. 3f). Its paralogue *UPP2* is not expressed in this tumour type (Extended Data Fig. 6a), as observed in The Cancer Genome Atlas (TCGA) data. We also observed from the Human Protein Atlas public database that *UPP1* gene expression is extremely low in normal pancreas and UPP1 protein expression is high in PDA (Extended Data Fig. 6b,c). Consistently, *UPP1* expression is also high in several other non-PDA cancers from TCGA, with colon and prostate cancers being the notable exceptions (Extended Data Fig. 6d). High *UPP1* also predicted poor survival outcome in lung, stomach, liver and renal cancers (Extended Data Fig. 6e). Indeed, we tested uridine utilization in several non-PDA cancer cell lines and observed a modest uridine-derived increase in metabolic activity (Extended Data Fig. 6f), supporting the potential relevance of this metabolite in other cancers.

To independently validate UPP1 expression in human PDA, we used patient samples to assay tumoural *UPP1* expression by RNAscope (Fig. 3g,h and Extended Data Fig. 7a), the cellular distribution of *UPP1* expression by single cell RNA sequencing (Extended Data Fig. 7b,c), and UPP1 expression by immunohistochemistry (Extended Data Fig. 7d). Collectively, these data were consistent with public databases, illustrating UPP1 upregulation in human PDA relative to normal pancreas tissue. In addition, our histological assessment of injured acinar cells indicated that UPP1 is upregulated upon pancreatic injury (Extended Data Fig. 7d), suggesting a potential role in PDA formation. Finally, stratification of tumour datasets also showed that the high expression of *UPP1* predicted poor overall survival outcome in three out of the four PDA patient cohorts we analysed (Fig. 3i and Extended Data Fig. 7e). Together, these data support a key role for UPP1 in PDA.

KRAS–MAPK pathway regulates UPP1

KRAS mutations are the signature transforming event observed in the majority of PDAs[31]. Using data from TCGA, we determined that PDA with the $KRAS^{G12D}$ mutation express higher levels of *UPP1* than those with no *KRAS* alteration (Fig. 3j). We also analysed the CCLE protein expression data of *KRAS* wild-type versus mutant cell lines across cancers and from lung or colorectal cancer. A targeted analysis of PDA lines was not performed because *KRAS* mutations are observed in all but one line in the CCLE. From the pan-analysis ($n = 374$), we observed a borderline, non-significant, association between *KRAS* status and UPP1 expression ($P = 0.09$) (Extended Data Fig. 8a), but lung cancer cell lines ($n = 79$) showed a significant association ($P = 0.003$). Colorectal cancer lines ($n = 30$) showed no difference or slightly reduced UPP1 in mutant *KRAS* lines (Extended Data Fig. 8a), consistent with the data in Extended Data Fig. 6d. The colon cancer results may be owing to the timing of *KRAS* mutation in tumour evolution and differences in its tissue-specific function relative to that of lung and PDA[32].

To experimentally test the role of mutant KRAS on UPP1 expression, we first queried published microarray data from our doxycycline-inducible KRAS (iKras) mouse model of PDA[33,34]. Mutant KRAS promoted *Upp1* expression in a subcutaneous xenograft model in vivo and iKras PDA cell lines in vitro (Extended Data Fig. 8b). Consistently, in vitro validation experiments in two additional, independent iKras cell lines confirmed doxycycline (via KRAS) induction of *Upp1* (Fig. 3k,l).

We previously showed that KRAS-mediated regulation of anabolic glucose metabolism in PDA occurs via mitogen-activated protein kinase (MAPK) signalling and MYC-dependent transcription[34]. In human and mouse cell lines, the pharmacological inhibition of MAPK reduced UPP1 transcript and protein, concurrent with the suppression of pERK (Fig. 3m,n and Extended Data Fig. 8c–j). MAPK inhibition also blocked the catabolism of uridine, as reflected by intracellular uridine accumulation, changes in a spectrum of other metabolites (Extended Data Fig. 8k), and a suppressed uridine-fuelled proliferation (Fig. 3o and Extended Data Fig. 8l). In contrast to the previous KRAS mechanism[34], MYC inhibition did not alter *UPP1* expression or

appear among the transcription factors binding to the *UPP1* promoter (Extended Data Fig. 9a–c), suggesting that MYC does not mediate KRAS regulation of UPP1 in PDA.

Nutrient availability modulates UPP1

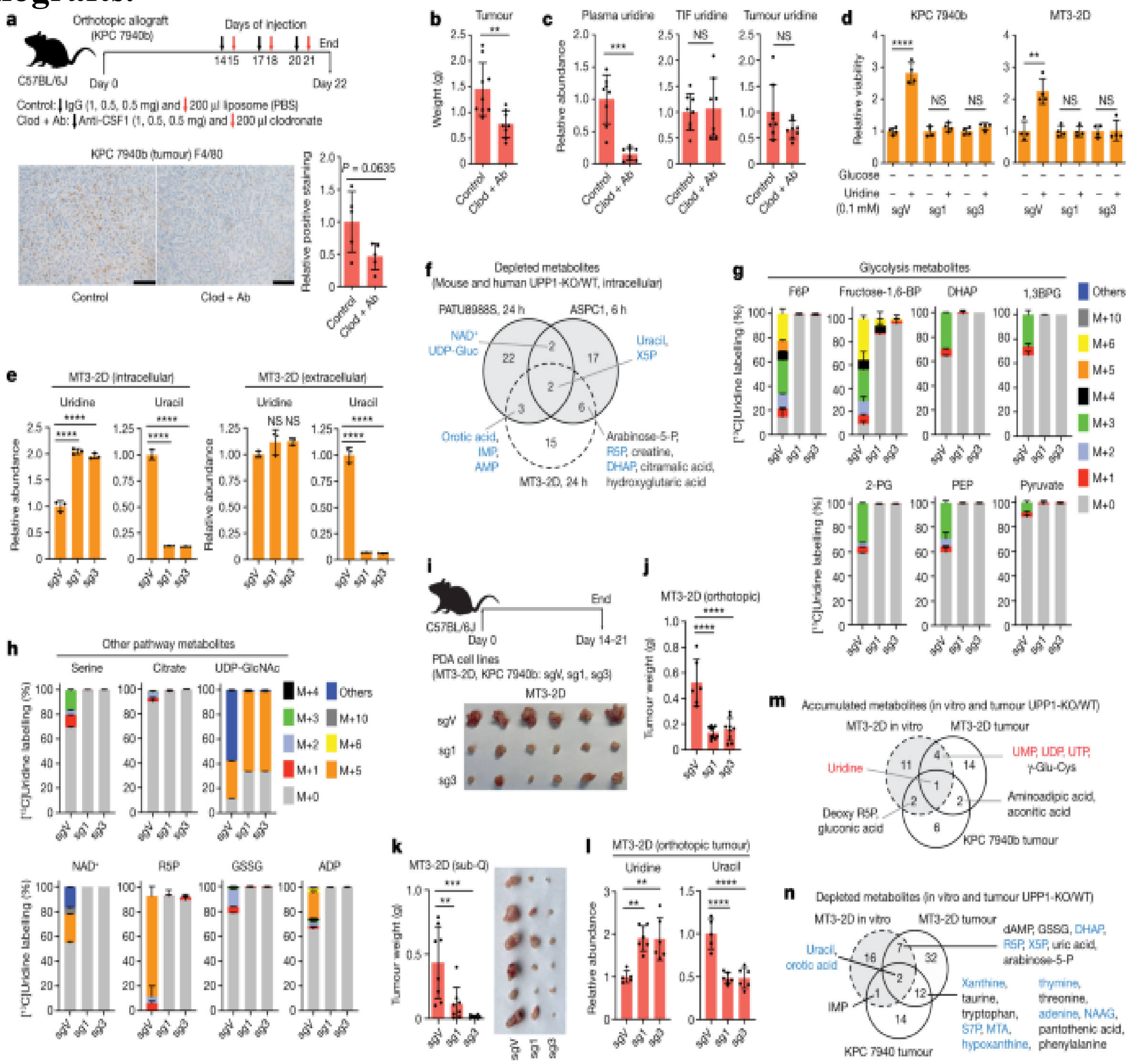
Given that glucose availability influences the use of uridine-derived ribose, we hypothesized that a glucose-depleted microenvironment triggers PDA to upregulate UPP1 as a compensatory response. Indeed, the removal or reduction of glucose in the medium induced a strong increase in UPP1 expression (Fig. 3m–n and Extended Data Figs. 8c–e,h–j and 9d), which was attenuated in uridine-supplemented medium. Consistently, we found that *Upp1* was strongly induced in a mouse tumour-derived cell line from the KPC model (*p48-cre;LSL-Kras*^{G12D}*;LSL-Trp53*^{R172H}) when cultured with TIF medium[30] or implanted as orthotopic allografts, relative to cells cultured in routine medium with high glucose (Extended Data Fig. 9e). Thus, KRAS–MAPK signalling and a nutrient-deprived TME may both be responsible for high UPP1 expression.

UPP1-KO blunts PDA tumour growth

Uridine concentration is reported to be around twofold higher in TIF than in plasma[30]. In looking for a cellular source supplying uridine to tumours, we observed from our previously generated dataset[14] that macrophages release uridine and uracil in vitro when differentiated and polarized to a tumour-educated fate with PDA-conditioned medium (Extended Data Fig. 10a). Thus, we tested the role of tumour-associated macrophages (TAMs) in supplying intratumoural uridine by depleting macrophages from mouse allograft tumours. To this end, we treated mice with a CSF1 antibody and clodronate liposome combination (Fig. 4a), which depletes TAMs and suppressed orthotopic tumour growth[35,36] (Fig. 4a,b). We observed a reduction in the plasma uridine level by around eightfold upon macrophage depletion, concomitant with an increased plasma uracil level. However, uridine and uracil levels in the TIF and tumour were not altered (Fig. 4c and Extended Data Fig. 10b). This marked effect on plasma uridine levels

following macrophage depletion indicates that macrophages may be important mediators of uridine production and/or release.

Fig. 4: UPP1-KO impairs the growth of orthotopic pancreatic tumour allografts.



a, Schematic of macrophage depletion (top) and validation by immunohistochemistry with F4/80 monoclonal antibody (bottom). Scale bars, 100 µm. Clod + Ab denotes anti-mouse CSF1 antibody and clodronate liposome. **b**, Tumour weight from control and macrophage-depleted mice at end point. **c**, Relative plasma, TIF and whole-tumour uridine levels in samples from the control and macrophage-depleted groups, measured by LC–MS. **d**, CellTiter Glo assay indicating relative viability (via ATP) of UPP1-KO (sg1 and sg3) and non-targeting control vector (sgV) mouse PDA

cells. **e**, Relative intracellular and extracellular uridine and uracil levels in the control and UPP1-KO MT3-2D mouse PDA cell lines after 1 mM uridine supplementation, as determined by LC–MS. **f**, Venn diagram showing metabolites depleted in vitro upon UPP1-KO in the human (PATU8988S and ASPC1) and mouse (MT3-2D) PDA cell lines, determined by LC–MS. Metabolites in blue are those from the glycolysis, nucleotide biosynthesis or pentose phosphate pathways. Arabinose-5-P, arabinose-5-phosphate; IMP, inosine monophosphate; UDP-Gluc, uridine diphosphate-glucose; X5P, xylulose 5-phosphate. **g**,**h**, Stable isotope tracing showing mass isotopologue distribution of 1 mM [$^{13}C_5$]uridine-derived carbon in metabolites from glycolysis (**g**) and other pathways (**h**) in the MT3-2D cells after 6 h culture. 1,3BPG, 1,3-bisphosphoglcerate; GSSG, oxidized glutathione. **i**, Top, schematic of tumour studies. Bottom, representative photograph of tumours collected from mice orthotopically implanted with the control vector (sgV) or UPP1-KO (sg1, sg3) MT3-2D cells. **j**, Tumour weight at end point of C57BL/6J mice orthotopically implanted with the MT3-2D cells. **k**, Weight and photograph of tumours collected at the end point after subcutaneous implantation of MT3-2D cells into the left and right flanks of C57BL/6J mice. **l**, Relative tumour uridine and uracil abundance in the orthotopic tumour samples described in **i**,**j**, as determined by LC–MS. **m**,**n**, Venn diagram showing metabolites commonly accumulated (**m**) and depleted (**n**) in UPP1-KO tumours implanted in mice and UPP1-KO cultured cells versus the wild type, as determined by LC–MS. **m**, Metabolites in red are associated with nucleotide biosynthesis. **n**, Metabolites in blue are associated with glycolysis, nucleotide biosynthesis or PPP. MTA, 5′-methylthioadenosine; NAAG, *N*-acetylaspartylglutamic acid; S7P, sedoheptulose 7-phosphate. See Methods ‘Statistics and reproducibility’ section for additional information.

To determine whether UPP1 supports tumour growth in vivo, we generated two independent models of UPP1-KO in the syngeneic mouse pancreatic cancer lines MT3-2D and KPC 7940b. Two single guide RNA (sgRNA) constructs targeting *Upp1* (sg1 and sg3) were compared to a vector control (sgV). We confirmed that glucose-deprived mouse UPP1-KO cells were not rescued by uridine in vitro (Fig. 4d). Furthermore, metabolomics profiling confirmed intracellular uridine accumulation and a reduced intracellular and extracellular uracil (Fig. 4e)—consistent with a blocked uridine catabolism.

These were accompanied by a change in a spectrum of other metabolites that were also altered in the human UPP1-KO cell lines (Fig. 4f and Extended Data Fig. 10c). Finally, isotope tracing of uridine ribose-derived carbon illustrated conclusively that UPP1-KO in the mouse lines restricted the use of uridine to fuel central carbon metabolism (Fig. 4g,h). This battery of metabolic assays confirmed a successful UPP1-KO in lieu of a mouse UPP1 antibody. Notably, the UPP1-KO cells did not differ from controls in terms of proliferation under optimal culture where glucose is present (Extended Data Fig. 10d).

We orthotopically implanted these two mouse lines and the corresponding UPP1-KO lines into the pancreas of syngeneic hosts and assessed tumour weight at end point (Fig. 4i). In both lines, contrary to the lack of a proliferative defect in vitro, we observed a markedly reduced tumour growth following UPP1-KO (Fig. 4j and Extended Data Fig. 10e). Similar results were also reproduced in an immunocompetent subcutaneous model using the MT3-2D cell lines (Fig. 4k). Metabolomic profiling of the orthotopic tumours revealed an increase in tumoural uridine and a drop in uracil in the UPP1-KO tumours as well as a profound change in the metabolome relative to vector controls (Fig. 4l–n and Extended Data Fig. 10f,g). In addition, compared with in vitro data, we observed the accumulation of a similar collection of metabolites (Fig. 4m) and a strong depletion of uracil, components of PPP, glycolysis and nucleotide metabolism (Fig. 4n).

The marked anti-tumour effect of UPP1-KO prompted us to look for changes in the in vivo microenvironment. Histological analysis of F4/80 staining revealed no differences in macrophage content between UPP1-KO and vector control tumours. However, the UPP1-KO tumours had lower vessel density (CD31) and more anti-tumour T cell infiltration (CD8 T cells) (Extended Data Fig. 10h). Together, our data indicate that UPP1 and uridine are important in PDA growth. The contrast between the in vitro and in vivo growth phenotype further highlights the role of nutrient availability, as well as the potential involvement of the TME and immune cell subsets in influencing the necessity of UPP1 in vivo.

Discussion

The metabolic features of PDA drive disease aggression and therapeutic resistance and present new opportunities for therapy[2,6]. Despite this, the range of nutrients used by PDA cells is poorly understood. We addressed this problem by applying high-throughput in vitro nutrient screening and found that under glucose-restricted conditions and KRAS–MAPK signalling activation, uridine serves as a nutrient source for PDA cells. This aligns with previous studies, where uridine rescued glucose deprivation-induced stress in astrocytes and neurons[20,21,23]. Indeed, others have shown that the uridine-mediated rescue is UPP1-dependent and induced by glucose availability, and it was proposed that UPP1 functions in this capacity to support bioenergetics by providing nucleotides[20]. We show that the uridine ribose ring fuels both energetic and anabolic metabolism in PDA cells. We also found that in addition to KRAS–MAPK, uridine utilization axis is regulated by a yet unknown rheostat sensing the upstream availability of glucose and/or uridine. These are newly identified regulators of UPP1, adding to p53 regulation[37]. Exploration of these regulatory pathways in PDA and other cancers may hold translational promise.

Nutrients in the TME can be derived from serum or the various cell types that make up the tumour. Although discovery methods involving conditioned medium and metabolomics are high-throughput, they tend not to capture the complex metabolic interactions of the TME. Biolog assays provide an unbiased approach to assess metabolic fuel utilization. Here we used this system to first obtain source-agnostic information about the nutrients utilized by PDA cells, before conducting a targeted analysis of potential sources providing exogenous uridine to tumours. We observed evidence of uridine enrichment in the TIF of mouse pancreatic tumours, tumour consumption of plasma-derived uridine by in vivo isotope tracing, reduction of the plasma uridine pool upon whole-body macrophage depletion, and in vitro micromolar release of uridine from naive macrophages and TAMs. Together, these findings illustrate the complexity of nutrient availability in the TME and suggest a model in which cells inside and outside PDA tumours fuel cancer metabolism with uridine.

There is a growing appreciation for the importance of nucleosides in cancer, including inosine, thymidine and deoxycytidine[6,14,38,39]. Inosine is consumed in melanomas by both cancer and $CD8^+$ effector T cells[40]. The

upregulation of nucleoside usage under nutrient deprivation[40,41,42], especially in immune and PDA cells, supports the idea that metabolic competition contributes to immunosuppression and tumour progression. Along these lines, we show that UPP1-KO in an orthotopic syngeneic model of PDA severely blunts tumour growth, thus the UPP1–uridine scavenging axis is important for PDA cells. RNA is another important source of uridine for glucose-starved cells and may be relevant for PDA cells, which readily scavenge intracellular (that is, autophagy) and extracellular biomolecules to fuel metabolism[5]. There are also several non-tumoural cell types that utilize uridine for various purposes[20,21,23,43], and this complexity is a promising area for future study. Collectively, our data identify the uridine–UPP1 axis as a driver of compensatory metabolism and support the therapeutic tractability of nucleoside metabolism in solid tumours.

Methods

Cell culture

The PDA cell lines A549, HT1080, HCT116 and U2OS and human pancreatic nestin-expressing cells were purchased from the American Type Culture Collection (ATCC) or the German Collection of Microorganisms (DSMZ). The human pancreatic stellate cells (hPSC) and mouse PDA cell lines KPC 7940b and MT3-2D were provided under a material transfer agreement by R. Hwang, G. Beatty and D. Tuveson, respectively. iKras cell lines A9993 and iKRAS 9805 were derived as described[33]. The identity of cell lines was confirmed by STR profiling, and lines were routinely tested for mycoplasma using MycoAlert (Lonza, LT07-318). For routine propagation, unless otherwise indicated, all cell lines were cultured in high-glucose DMEM (Gibco, 11965092) supplemented with 10% FBS (Corning, 35-010-CV) at 37 °C and 5% CO_2. PBS (Gibco, 10010023) was used for cell washing steps unless otherwise indicated. For treatments, the following inhibitors were used: MYC, fedratinib (MedChemExpress, HY-10409) and 10058-F4 (Cayman Chemical, 15929); MEK1, trametinib (Selleckchem, S2673).

Biolog metabolic assay

In the initial phenotypic screen, the 22 cell lines were grown in 96-well PM-M1 and PM-M2 plates (Biolog, 13101 and 13102). The assay was set up such that one well was used per test metabolite substrate, accompanied by three replicates of positive (glucose) and negative (blank) control wells. The RMA from substrate catabolism in the cells was measured using Biolog Redox Dye Mix MB. In brief, the cell lines were counted, and their viability assessed using Trypan Blue Dye (Invitrogen, T10282). The cells were then washed two times with Biolog Inoculating fluid IF-M1 (Biolog, 72301) to remove residual culture medium. Then, a cell suspension containing 20,000 cells per 50 µl was prepared in Biolog IF-M1 containing 0.3 mM glutamine and 5% dialysed FBS (dFBS) (Hyclone GE Life Sciences, SH30079.01) and plated into PM-M1 and PM-M2 96-well plates at 50 µl per well. Plates were incubated for 24 h at 37 °C and 5% CO_2, after which 10 µl Biolog Redox Dye Mix MB (Biolog, 74352) was added to each well. Plates were sealed to prevent the leakage of CO_2. The reduction of the dye over time was measured as absorbance (A590-A750) using the OmniLog PM-M instrument (Biolog, 93171) for 74.5 h at 15 min intervals. To account for proliferation or cell number in the Biolog screening assay, CyQUANT was used for normalization.

The data were processed and normalized using the opm package[45] version 1.3.77 in the R statistical programming tool. After removing CFPAC1 (atypically high signal across the plate), the maximum metabolic activity per cell line was taken as its main readout for substrate avidity (Extended Data Fig. 1a) and normalized by subtracting the median negative control signal for a given cell line from all other values for that cell line. Heat map visualization of the data was plotted using heatmap2 and ComplexHeatmap packages in R.

Correlation of Biolog metabolites to gene expression of enzymes

High-confidence metabolites from the Biolog screening assay were correlated (Spearman Correlation) to gene expression data for enzymes associated with metabolite usage. Genes with high correlation co-efficient to a given metabolite were chosen for further analysis.

NADH assay

Cells were seeded in 96-well plates at 10,000 cells per well directly into the indicated medium conditions. Following the incubation period at 37 °C and 5% CO_2 (24, 48 or 72 h), MTT (Thermo Scientific, L1193903) was added directly to the wells containing medium. Cells were incubated at 37 °C and 5% CO_2 for 1 h, after which the medium and MTT reagent were carefully removed. Next, to each well, 50 µl of DMSO (Sigma-Aldrich, D2650) was added followed by 5 min incubation at room temperature before measuring absorbance at 570 nm.

CyQUANT proliferation assay

Cells were seeded at 20,000 cells per well for the screening study or 2,000 cells per well for proliferation assays in growth medium in 96-well plates (Corning, 3603). For proliferation assays, the culture medium was removed the next day, followed by a gentle 1× wash with PBS. Treatment medium was then applied, and the cells incubated at 37 °C and 5% CO_2 until they reached ~70% confluence. Medium was then carefully aspirated, and the plate with cells attached was moved to −80 °C for at least 24 h to ensure complete cell lysis. To prepare the lysis buffer and DNA dye, CyQUANT Cell Lysis Buffer and CyQUANT GR dye (Invitrogen, C7026) were diluted in water at 1:20 and 1:400, respectively. The frozen cells were then thawed and 100 µl of the lysis buffer was added to each well. Thereafter the plate was covered to prevent light from inactivating the GR dye and was placed on an orbital shaker for 5 min before measurement. Fluorescence from each well, indicating GR dye binding to DNA, was then measured utilizing a SpectraMax M3 Microplate Reader with SoftMax Pro 5.4.2 software at an excitation wavelength of 480 nm and an emission wavelength of 520 nm.

ATP-based viability assay

Cells were seeded in quadruplicates at a density of 2,500–5000 cells in 100 µl DMEM per well of the white walled 96-well plates (Corning/Costar, 3917). Next day, the medium was aspirated, each well was washed with 200 µl PBS after which treatment medium was introduced. At the end of the experiment duration (48 or 72 h), relative proliferation was determined with CellTiter-Glo 2.0 Cell Viability Assay Kit (Promega, G9243) and the luminescence quantified using a SpectraMax M3 Microplate Reader.

Live-cell proliferation assay

Cells were seeded in 96-well plates at 1,000 cells per well in 100 µl of growth medium and incubated overnight at 37 °C, 5% CO2. After 24 h, medium was changed, and the cells were incubated for a further 72 h during which cell proliferation was determined by live-cell imaging on a BioSpa Cytation.

UPP1 CRISPR–Cas9 knockout

The expression vector pspCas9(BB)-2A-Puro (PX459) used to generate the UPP1 CRISPR–Cas9 constructs was obtained from Addgene (Plasmid 48139). The plasmid was cut using the restriction enzyme BbsI followed by the insertion of human or mouse uridine phosphorylase 1 sgRNA sequences (Supplementary Table 4), as previously described[46]. The human and mouse sequences were obtained from the Genome-Scale CRISPR Knock-Out (GeCKO) library. For transfection, the human or mouse PDA cells were seeded at 150,000 cells per well in a 6-well plate a day earlier. The cells were transfected with 1 µg of plasmid pSpCas9-UPP1 using Lipofectamine 3000 Reagent (Invitrogen, L3000001) according to manufacturer's instruction. After 24 h, the selection of successfully transfected cells was commenced by culturing the cells with 0.3 mg ml^{-1} puromycin in DMEM. The puromycin-containing medium was replaced every two days until selection was complete, as indicated by the death and detachment of all non-transfected cells. Thereafter, the successfully transfected cell lines were expanded and clonally selected after serial dilution.

siRNA experiments

Approximately 5×10^5 ASPC1 cells were seeded per 6-cm dish for 24 h in the growth medium. On day 2, medium was changed and the respective SMARTpool siRNA-containing medium was added. siRNA transfection was performed using Lipofectamine RNAiMAX (ThermoFischer Scientific, 13778075) according to the manufacturer's instructions. For the transfection, Opti-MEM Reduced Serum Medium (31985-062) was used and siRNAs were added at a concentration of 20 nM. Cells were transfected for 48 h after which the cells were trypsinized, counted, and plated for MTT assay. The

remaining cells were pelleted, and RNA was extracted for qPCR. The siRNAs used were as follows: non-targeting control (D-001810-01-05), SMARTpool ON-TARGETplus Human PGM2 (55276) (L-020785-01-0005), ON-TARGETplus Human UCK1 (83549) (L-004062-00-0005) and ON-TARGETplus Human UCK2 (7371) (L-005077-00-0005).

Reverse transcription with qPCR

Cells were seeded at a density of 5×10^5–8×10^5 cells per well, allowed to attach overnight, treated where applicable and pelleted after 24 h. RNA samples were isolated using the RNEasy Plus Mini Kit (QIAGEN, 74134) according to the manufacturer's instructions. RNA purity was assessed using a NanoDrop One (ThermoFisher Scientific, ND-ONE-W). Thereafter, 1 µg of the RNA samples were reverse transcribed to cDNA using the iScript cDNA Synthesis Kit (Bio-Rad, 1708890) according to the accompanying instructions. qPCR was performed on the QuantStudio 3 Real-Time PCR System (ThermoFisher Scientific, A28131) or Applied Biosystems StepOne Plus instrument (software version 2.3) using Power SYBR Green PCR Master Mix (ThermoFisher Scientific, 4367659). The reactions were run at 10 µl total volume consisting of 5 µl SYBR, 2 µl nuclease free water, 2 µl of cDNA after diluting 1:4 in water, 0.5 µl of 10 µM forward (F) primer, and 0.5 µl 10 µM reverse (R) primer. Primer sequences are listed in Supplementary Table 5. The gene expression was calculated as ΔC_t and *RPS21* or *ACTB* was used as a housekeeping gene.

Western blotting

Following culture, medium was aspirated, and the wells washed one time with PBS. Thereafter, 100 µl of radioimmunoprecipitation assay buffer (Sigma-Aldrich, R0278) to which phosphatase and protease inhibitors were added, as transferred to each well to lyse the cells. Lysis and the collection of the lysates were completed on ice. Following a 5- to 10-minute incubation on ice, lysates were collected into 1.5 ml Eppendorf tubes and centrifuged at 4 °C for 10 min at 18,000*g* to extract the sample supernatant. Protein concentration of the samples for western blot analysis were measured using Pierce BCA Protein Assay Kit (ThermoFisher, 23227) according to the manufacturer's instructions. For the running step, samples were loaded at

20–25 µg protein per lane along with the SeeBlue Plus2 protein ladder (Invitrogen, LC5925) and run at 120 V on an Invitrogen NuPAGE 4–12% Bis-Tris gel (ThermoFisher, NP0336BOX). Thereafter, the separated proteins were transferred to methanol-activated PVDF membranes (Millipore) at 25 V for 1 h. Following this, membranes were immersed in blocking buffer (5% blotting-grade blocker (Bio-Rad, 1706404) in TBS-T solution: tris-buffered saline (Bio-Rad, 1706435) with 0.1% Tween-20 (Sigma-Aldrich, 9005-64-5)) for ~1 h on a plate rocker at room temperature. Next, membranes were washed 3 times with TBS-T at 10 min per wash, immersed in the indicated primary antibodies, and incubated overnight at 4 °C on a plate rocker. The antibodies used were diluted in blocking buffer at dilutions recommended by the manufacturer. The following day, the primary antibody was removed, and the membrane was washed 3 times with TBS-T and on a plate rocker for 5 min per wash. Immediately after, the membrane was incubated for 1 h and with gentle rocking at room temperature in the appropriate secondary antibody diluted 1:10,000 in TBS-T. Lastly, the membrane was washed 3 times in TBS-T at 10 min per wash and incubated in chemiluminescence reagent (Clarity Max Western ECL Substrate, 705062) according to the manufacturer's instructions. Subsequently, blot images were acquired on a Bio-Rad ChemiDoc Imaging System (Image Lab Touch Software version 2.4.0.03). The following primary antibodies were used in this study and at 1:1,000 dilution: anti-UPP1 (Sigma-Aldrich, HPA055394), anti-c-MYC (Cell Signaling, 5605S), anti-pERK (Cell Signaling, 9106L), anti-ERK (Cell Signaling, 9102S), anti-PGM2 (Invitrogen, PA5-31378), and anti-Vinculin (Cell Signaling, 13901S). The following secondary antibodies were used: anti-rabbit-HRP (Cell Signaling, 7074S), and anti-mouse-HRP (Cell Signaling, 7076P2). The uncropped, unprocessed images of the western blots are presented in Supplementary Figs. 1–9.

Mouse tumour studies

Animal studies were performed at the University of Michigan, the Institute of Cancer Research (ICR), and the University of Chicago according to approved protocols. Specifically, for studies at the University of Michigan, the Institutional Animal Care and Use Committee (IACUC) PRO00010606 was followed; Institute of Cancer Research studies conformed to UK Home

Office Regulations under the Animals Scientific Procedures Act 1986 and national guidelines (project licence P0A54750A protocol 5); University of Chicago IACUC protocol 72587 was followed. Mice were housed in a pathogen-free animal facility at a maximum of five mice per cage with a 12 h light/12 h dark cycle, 30–70% humidity and 20–23 °C temperatures. Mice were provided water and fed ad libitum with chow (5L0D – PicoLab Laboratory Rodent Diet).

Pancreatic tumour models

For mouse studies at the University of Michigan, male and female 6- to 8-week-old C57BL/6J mice were obtained from The Jackson Laboratory (strain 000664) and maintained in the facilities of the Unit for Laboratory Animal Medicine (ULAM) under specific pathogen-free conditions. Prior to tumour cell injection, wild-type or UPP1-KO mouse cell lines (either derived from MT3-2D or KPC 7940b) were collected from culture plates according to standard cell culture procedures. The cells were counted, washed once with PBS and resuspended in 1:1 solution of serum-free DMEM and Matrigel (Corning, 354234). For the orthotopic surgical procedure, mice were anaesthetized using inhalation anaesthesia. The surgical site was sterilized by swapping with iodine (Povidine-Iodine Prep Pad, PDI, B40600). This was followed by incision on the left flank using sterilized instruments.Thereafter, the cell lines were injected into the pancreas and the incision sutured. Cell injection was as follows: 50,000 or 100,000 cells in 50 μl final volume for orthotopic implantation or ~1×10^6 cells in 100 μl final volume for subcutaneous implantation. Animals were monitored regularly and all orthotopic experiments were concluded ~3–4 weeks after injection.

For the studies at the Institute of Cancer Research, female ~6-week-old C57BL/6NCrl mice were purchased from Charles River Laboratories (strain 027). Prior to tumour cell injection, MT3-2D sgVector (sgV) and the UPP1-KO cells (sg1 and sg3) were trypsinized according to standard cell culture protocols. Cells were washed with PBS and resuspended in 1:1 Hank's balanced salt solution (HBSS; Gibco, 14025092) and Matrigel (Corning, 354234). Following surgical incision, 50,000 cells in 20 μl final volume were injected into the pancreas. Tumour growth was monitored by palpating

three times per week. Studies were terminated when the animals injected with the vector reached a high tumour burden based on palpation.

For the studies at the University of Chicago, C57BL/6J mice 8–12 weeks of age were purchased from Jackson Laboratories (strain 000664). About 2.5 × 10^5 cells per tumour were resuspended in 20 µl of 5.6 mg ml^{-1} Cultrex Reduced Growth Factor Basement Membrane Extract (RGF BME; R&D Biosystems, 3433-010-01) and serum-free RPMI (SF-RPMI) solution. The basement membrane extract–cell mixture was injected into the splenic lobe of the pancreas of the mice, as previously described[43]. After implantation, end point was determined by abdominal palpation and daily monitoring of body weight.

Macrophage depletion

The KPC 7940b cell line was orthotopically implanted into approximately 8-week-old male and female C57BL/6J mice, as above. Two weeks after injection, mice were randomized into two groups: control or macrophage depletion. The control group mice were treated on day 1 with 1 mg IgG (InVivoMAb rat IgG1 Isotype control, anti-trinitrophenol, BE020, Bio X Cell) and on day 2 with 200 µl Control Liposome (PBS) (CP-005-005, Liposoma). The macrophage depletion mice were treated on day 1 with InVivoMAb anti-mouse CSF1 (anti-CSF1, BE0204, Bio X Cell) and day 2 with 200 µl clodronate liposomes (CP-005-005, Liposoma). Two subsequent treatment sequences were administered at 2-day intervals as follows: 0.5 mg IgG or anti-CSF1 followed the next day by 200 µl control liposome or clodronate liposomes for the control and depletion groups, respectively. The experiment was terminated after one week. From each mouse, blood samples were collected into EDTA BD Vacutainer K2 EDTA 3.6 mg (36784) and centrifuged at 200*g* for 5 min for plasma collection. In addition, tumours were collected, weighed and used for the extraction of TIF, as below.

TIF collection

TIF was isolated from tumours as described[30]. In brief, tumours were rapidly dissected after euthanizing the animals. Tumours were weighed and rinsed in blood bank saline solution (150 mM NaCl) and blotted on filter paper

(VWR, 28298–020) until dry. Tumour isolation was done in less than 3 min to minimize the time the tumour was ischaemic prior to TIF isolation. Tumours were cut in half and put onto 20-µm nylon mesh filters (Spectrum Labs, 148134) on top of 50 ml conical tubes, and centrifuged for 10 min at 4 °C at 400*g*. TIF was then processed for metabolomics in a similar manner as plasma, as described below.

In vivo delivery of isotopically labelled uridine

Uridine-derived ribose carbon was traced in vivo using [$^{13}C_5$]ribose-labelled uridine (Cambridge Isotope Laboratories, CLM-3680-PK). Specifically, mice bearing orthotopic or subcutaneous tumours were generated, as described above, using KPC 7940b cell lines. For the orthotopic tumours, after the tumours became palpable (3 weeks after implantation), mice were injected intraperitoneally with 200 µl of 0.2 M [$^{13}C_5$]uridine. For the subcutaneous models, 50 µl of 0.2 M [$^{13}C_5$]uridine was injected directly into the tumours 3 weeks after implantation. Tumours were collected 1 h after uridine injection and processed for isotope tracing, as detailed below.

Mass spectrometry-based metabolomics

Metabolomics sample preparation

For in vitro extracellular (medium) and intracellular metabolomic profiling, PDA cells were seeded in triplicates in a 6-well plate at 4–6 × 10^6 cells per well in growth medium. A parallel plate for protein estimation and sample normalization was also set up. After overnight incubation, the culture medium was aspirated and replaced with medium containing treatments or supplemented metabolites of interest. The cells were then cultured for a further 24 h. Thereafter, for extracellular metabolites, 200 µl of medium was collected from each well into a 1.5 ml Eppendorf tube and to that 800 µl ice-cold methanol was added. For intracellular metabolites, the remaining medium was aspirated, and samples washed once with 1 ml PBS before incubation with 1 ml ice-cold 80% methanol on dry ice for 10 min. Thereafter, cell lysates were collected from each well and transferred into separate 1.5 ml Eppendorf tubes. The samples were then centrifuged at

12,000*g*. For each experimental condition, the volume of supernatant to collect for drying with SpeedVac Vacuum Concentrator (model: SPD1030) was determined based on the protein concentration of the parallel plate.

For tumours, the samples were flash frozen in liquid nitrogen upon collection. Tumours of approximately equal weight (<100 mg) were collected per sample per experimental group. The tumours were then put into 2 ml Eppendorf tubes to which 1 ml of ice-cold 80% methanol (diluted in 20% H_2O). Metallic beads were added to each tube and samples were shaken and homogenized on an Retsch TissueLyser II (129251128) in intervals of 30 s until fully homogenized. Samples were then centrifuged at 12,000*g* and supernatant collected for further processing.

Targeted metabolomics

The collected supernatants were dried using SpeedVac Vacuum Concentrator, reconstituted with 50% v/v methanol in water, and analysed by LC–MS, as described in detail previously[47]. Data were analysed with Agilent Masshunter Workstation Quantitative Analysis for QQQ version 10.1, build 10.1.733.0.

Stable isotope tracing

For stable isotope tracing in cells, [$^{13}C_5$]uridine (Cambridge Isotope Laboratories, CLM-3680-PK) was supplemented at 0.1 mM or 1 mM for in vitro assays. In brief, wild-type or UPP1-KO cells were cultured overnight in regular medium. Next day, cells were washed once followed by the introduction of medium containing the indicated amounts of glucose, dialysed FBS and supplemented with labelled uridine. In parallel, a similar experiment was set up for unlabelled uridine. The cell lines were then cultured in the uridine-supplemented medium for 24 h or as otherwise indicated, followed by sample collection, as detailed for unlabelled intracellular metabolomics above. Labelled tumours were similarly collected as detailed above. Samples were prepared for time-of-flight mass spectrometry, as described in detail previously[47], and analysed with Agilent MassHunter Workstation Profinder version 10.0, build 10.0.10062.0.

For the experiments were glucose and uridine concentrations were varied (5 and 0.1 mM and 1 and 0.1 mM, respectively) followed by stable isotope tracing, the cells were seeded at a density of 500,000 and treated with [$^{13}C_5$]uridine or unlabelled uridine in DMEM supplemented with dialysed FBS and the indicated concentration of glucose for 24 h. Then cells were washed with 1 ml cold PBS followed by the addition of 1 ml 2:2:1 methanol:acetonitrile:water at −20 °C to the wells on dry ice for 10 min. Cells were scraped from the dish. Next, samples were subjected to 3 cycles of 30 s vortex, 1 min liquid N_2, and 10 min 25 °C bath sonicate. Samples were then stored at −20 °C overnight and centrifuged at 14,000*g* at 4 °C for 10 min. A volume of 860 µl of the supernatant was transferred to a new tube and dried by SpeedVac Vacuum Concentrator. Protein pellets were also dried similarly to remove excess supernatant, resuspended in 400 µl 100 mM NaOH through repeated vortexing, 5 min incubation at 95 °C, and protein quantified by BCA assay (ThermoFisher, 23227). Dried supernatant pellets were resuspended in 2:1 acetonitrile:water at 1 µl per 2.5 µg protein and subjected to 2 cycles of: 5 min 25 °C bath sonicate, 1 min vortex. Samples were incubated at 4 °C overnight, then centrifuged at 14,000*g* at 4 °C for 10 min and the supernatant transferred to liquid chromatography vials and stored at −80 °C until analysis. For sample analysis, 4 µl of metabolite extracts were run on an Agilent 6545 Q-TOF Mass Spectrometer and an Agilent 1290 Infinity II LC system using a iHILIC-(P) Classic 2.1 mm × 100 mm, 5 µm column (HILICON, 160.102.0520) with iHILIC-(P) Classic Guard column (HILICON, 160.122.0520) attached. A column temperature of 45 °C and a flow rate of 250 µl min^{-1} was used. Mobile phases were A: 95% water, 5% acetonitrile, 20 mM ammonium bicarbonate, 0.1% ammonium hydroxide solution (25% in water), 2.5 µM medronic acid and B: 85% acetonitrile, 5% water, 2.5 µM medronic acid. Each sample was subjected to a linear gradient: 0–1 min 90% B, 1–12 min 35% B, 12–12.5 min 25% B, 12.5–14.5 min 25% B, 14.5–15 min 90% B, which was then followed by 4 min at 400 µl min^{-1} and 2 min at 250 µl min^{-1} at 90% B for re-equilibration. Chromatograms for selected metabolites were extracted in Skyline Daily (software version 22.2.1.256) and manually integrated according to an in-house list of standard *m*/*z* and retention times. Natural isotope abundance correction was performed, and peak areas plotted.

Quantification of TIF metabolite levels

For quantification of uridine and glucose in TIF, quantitative metabolite profiling of fluid samples was performed as previously described[30]. In brief, chemical standards were prepared and serially diluted in high-performance liquid chromatography grade water in a dilution series from 5 mM to 1 μM. Using the external standard library dilutions, we created a standard curve based on the linear relationship of the normalized peak area and the concentration of the metabolite. This standard curve was then used to interpolate the concentration of the metabolite in the TIF sample.

Clinical samples

Patients with pancreas resections for PDA from 2021 to 2022 at the University of Michigan Health System were included in the study. All haematoxylin and eosin (H&E)-stained slides were reviewed and diagnoses confirmed, and corresponding areas were carefully selected and marked. The collection of patient-derived tissues for histological analyses was approved by the Institutional Review Board at the University of Michigan (IRB number: HUM00098128). Tissues were fixed in 10% neutral buffered formalin and paraffin embedded using standard protocols before sectioning and staining.

Tissue microarrays

All specimens are from patients with pancreas resections for pancreatitis, cystic neoplasms, or PDA from 2002 to 2015 at the University of Michigan Health System. After fixation in 10% neutral buffered formalin (hours to a couple of days depending on the size of the tissue), samples were embedded in paraffin. All tissues were H&E stained, reviewed, and diagnoses confirmed. Corresponding areas were carefully selected and marked. Duplicate 1 mm diameter tissue cores from a total of 213 patient tissue samples were selectively punched and transferred to recipient tissue array blocks. Five tissue microarrays (TMAs) were set up and H&E and immunohistochemistry staining was performed on each TMA block using standard protocols. The TMA was previously published[48].

RNAscope

RNAscope was performed as previously described[49], and according to the manufacturer's protocol (ACD: 323100-USM). In brief, paraffin wax was removed with xylene and slides were rehydrated. Samples underwent antigen retrieval for 15 min. Samples were blocked for 30 min at room temperature with CoDetection Antibody Diluent and then incubated overnight at 4 °C with a primary antibody for panCK (Mouse anti-cytokeratin, pan reactive; 1:100; BioLegend; 628602) diluted in CoDetection Antibody Diluent. Protease digestion was performed for 13 min at room temperature. Human *UPP1* RNAscope probe (ACD; 509279; Lot: 21272A) was added to slides for 2 h at 40 °C. Samples were incubated with TSA-Cy3 fluorophore (1:2,000; Akoya Biosciences; NEL704A001KT) diluted in CoDetection Antibody Diluent. Following HRP blocking, slides were rinsed in PBS + 0.1% Tween-20 (PBST). Slides were stained with DAPI (1:30,000; Millipore Sigma; 10236276001) diluted in PBS for 15 min at room temperature. After rinsing in PBST, slides were incubated for 45 min at room temperature with secondary antibodies diluted 1:500 in CoDetection Antibody Diluent. Slides were rinsed with PBST and mounted in ProLong Gold Antifade Mountant (Invitrogen, P36930). Sections were visualized on a Leica SP5X upright confocal. For quantitation, 20× fields of view were imaged and analysed using FIJI/Image J (version 1.53c). For analysis, images were converted to 16 bit, the threshold was adjusted, and the area of *UPP1* expression was measured per 20× image.

Immunohistochemistry

Patients tissue slides were deparaffinized and rehydrated with graded Histo-Clear (National Diagnostics), ethanol, and water. Slides were quenched for 15 min in a methanol solution containing 1.5% hydrogen peroxide before antigen retrieval. Samples underwent antigen retrieval with sodium citrate buffer (2.94 g l^{-1} sodium citrate, 0.05% Tween-20, pH 6). Samples were blocked using blocking buffer (5% bovine serum albumin, 0.2% Triton X-100, in PBS) for 1 h at room temperature. After blocking, slides were incubated overnight at 4 °C with primary antibody (rabbit anti-UPP1; 1:200; Sigma-Aldrich, HPA055394) diluted in blocking buffer. Slides were rinsed in PBS and incubated for 1 h at room temperature with a biotinylated secondary antibody (horse anti-rabbit; 1:500; Vector Labs, BA-1100). After rinsing, slides were prepared for a colour reaction by incubating with

Vectastain Elite ABC Reagent (Vectastain Elite ABC-HRP Kit; Vector Labs, PK-6100) for 30 min at room temperature. Sections were developed with DAB (DAB Substrate Kit; Vector Labs, SK-4100) for 2 min, rinsed, and counterstained with haematoxylin. Slides were mounted in Permount Mounting Medium (Fisher). After drying, slides were imaged using an Olympus BX53F microscope, Olympus CP80 digital camera, and CellSens standard software.

Mouse tumours were fixed in 10% neutral buffered formalin for 48 h and embedded in paraffin as formalin-fixed paraffin-embedded (FFPE) blocks. Serial sections of 4 µm thickness were cut from FFPE blocks, deparaffinized in xylene, processed in graded alcohol, and rehydrated in water. One section was stained with H&E for histological analysis. The Dako Autostainer Link 48 automated immunostaining platform was used for all the below immunostainings. Anti-Cd31 monoclonal antibody (SZ31, DIA-310, Dianova) was used at a 1:75 dilution, followed by heat-induced epitope retrieval for 20 mins at 97 °C using a PT Link module (Agilent). Anti-Cd8 monoclonal antibody (clone 4SM15, 14-0808, eBioscience) was used at a 1:200 dilution, and anti-F4/80 monoclonal antibody (Clone-A3-1, MCA497G, Bio-Rad) was used at a 1:100 dilution. For these antibodies, EnVision FLEX Target Retrieval Solution (high pH; K800421-2, Agilent) and Nichirei anti-rat Histofine polymer reagent (41491F, Nichirei Biosciences) primary antibody detection kits were used. For CD3, polyclonal antibody (ab5690, Abcam) was used at a 1:400 dilution with EnVision FLEX Target Retrieval Solution (low pH of 6) for 20 mins in PT link module and detected with Vector Rabbit ImmPRESS HRP horse anti-rabbit IgG polymer kit (Vector Laboratories, MP-7401-50). Appropriate positive and negative controls were used in all runs. The Nanozoomer-XR C12000 (Hamamatsu) was used to scan whole stained sections. Antigen expression was scored using Definiens Test Studio Software (Definiens). F4/80 was quantified using Image J.

Immunohistochemistry of UPP1 expression in human normal and PDA tissues was also accessed from the Human Protein Atlas portal[50].

PDA dataset analysis

The human PDA microarray datasets with accession numbers GSE71729 (*n* = 46 normal pancreas vs 145 tumour tissues) and GSE62452 (*n* = 61 non-tumoural vs 69 tumour tissues) were obtained from NCBI GEO[51]. Differential gene expression between PDA and non-tumours were performed in R using the limma package (version 3.38.3). Kaplan–Meier overall survival (log-rank test) was performed after splitting the tumour samples per dataset into *UPP1*-high and *UPP1*-low subsets. For Kaplan–Meier analysis, human PDA tumour datasets and the accompanying clinical data from the following sources were used: GSE71729 (*n* = 145), TCGA data (*n* = 146), International Cancer Genome Consortium (ICGC, *n* = 267), and Puleo et al. (2018) (*n* = 288)[44]. The iKras mice data were obtained from NCBI GEO under the accession number GSE32277. TCGA expression data of tumours with *KRAS* wild type (*n* = 43) and $KRAS^{G12D}$ mutation (*n* = 42) were used to determine the relative expression of *UPP1* in $KRAS^{G12D}$ mutated tumours.

Pan-cancer dataset analysis

TCGA pan-cancer datasets including bladder, colon, oesophageal, lung, head and neck, prostate cancer and glioblastoma, were downloaded from Xena Platform from University of California Santa Cruz[52]. An additional colorectal dataset (GSE44076) was also used. For the comparisons, the normal or adjacent matched and unmatched normal samples were used. In total, 2,828 cancer tissue samples and 379 non-tumoural control tissue samples were analysed. These datasets were used to compare *UPP1* expression between cancer and non-cancer tissues.

CCLE gene analysis and PDA tumour data stratification

Gene expression data for uridine high and uridine low metabolizers were extracted from the CCLE (GSE36133). The subsets were then compared using the limma package in R to determine the differentially expressed genes in uridine-high metabolizers and/or consumers relative to the lower metabolizers and/or consumers. For the tumour stratification, samples in the dataset GSE71729 (*n* = 145) were ranked into *UPP1*-high and *UPP1*-low groups and compared as above to determine the genes differentially expressed in *UPP1*-high tumours. UPP1 protein expression analysis was

performed in *KRAS* mutant and wild-type cell lines using data from DepMap[53].

Pathway analyses

Pathway analyses were performed using DAVID functional annotation platform (https://david.ncifcrf.gov/, version 6.8) or gene set enrichment analysis (GSEA, version 4.0.3) with GSEAPreranked option. Ranking of genes was based on the product of the logFC and −log(*P* value). GSEA was run with default parameters, except gene set size filter set at min = 10. Gene ontology analyses were performed with DAVID.

Promoter analysis of *UPP1*

CiiDER[54] was used for predicting *UPP1* gene transcription factor sites. DNA sequence flanking the *UPP1* transcription start site (1,500 bases upstream, 500 bases downstream) was used to compare to JASPAR2020_CORE_vertebrates position frequency matrix model to generate a score of similarity. As transcription factor binding sites are variable and binding sites rarely match the model perfectly, a default deficit score of 0.15 was used, where deficit score of 0 represents prefect match. Top 10 transcription factors were obtained using the predicted UPP1-binding sites with respect to sequences from the human genome (GRCh38.94) and mouse genome (GRCm38.94).

Statistical analysis

Statistics were performed either with GraphPad Prism 8 (GraphPad Software Inc.) or using R version 3.5.2. Data from experimental groups were compared using the two-tailed *t*-test or analysis of variance (ANOVA) with post hoc corrections where applicable, and between biological (or in vitro) replicates. Data in all graphs represent the mean ± s.d. Statistical significance was accepted if $P < 0.05$. For data analysis and visualization in R, packages (with versions) used include dplyr (0.8.3), ggplot2 (3.3.5), gplots (3.0.1, heatmap.2 function), ComplexHeatmap (2.3.5), tidyverse (1.3.0) and VennDiagram (1.6.20).

Statistics and reproducibility

Figure 1. a, The use of >175 metabolites by 19 PDA cell lines and 2 non-PDA pancreatic cell lines was measured every 15 min for ~3 days (74.5 h) using the Biolog OmniLog device. The assay readout, RMA, was correlated with the expression level of metabolic genes in cell lines; human PDA data were used for subsequent analyses. Nutrient-deficient medium, no glucose, 0.3 mM glutamine and 5% dialysed FBS. c, $n = 4$ biologically independent samples per group. Statistical significance was measured by multiple unpaired two-tailed *t*-tests (two-stage step-up method) comparing RMA from cells in basal medium vs 1 mM uridine medium (both glucose-free), ****$P < 0.0001$. The experiments were performed twice with similar results. d, $n = 4$ biologically independent samples per cell line. The experiment was performed once.

Figure 2. a, $n = 4$ biologically independent samples per group per cell line. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. CAPAN2 (comparison between no glucose/uridine and no glucose + 1 mM uridine, ***$P = 0.0001$; no glucose/uridine and 1 mM glucose/no uridine, *$P = 0.021$; no glucose/uridine and 0.1 mM ribose, $P = 0.86$; no glucose/uridine and 1 mM ribose, **$P = 0.0093$; no glucose/uridine and 10 mM ribose, ****$P < 0.0001$). PATU8988S (comparison between no glucose/uridine and no glucose + 1 mM uridine, ****$P < 0.0001$; no glucose/uridine and 1 mM

glucose/no uridine, ****$P < 0.0001$; no glucose/uridine and 0.1 mM ribose, $P = 0.9817$; no glucose/uridine and 1 mM ribose, **$P = 0.0019$; no glucose/uridine and 10 mM ribose, ****$P < 0.0001$). DANG (comparison between no glucose/uridine and no glucose + 1 mM uridine, ****$P < 0.0001$; no glucose/uridine and 1 mM glucose/no uridine, ****$P < 0.0001$; no glucose/uridine and 0.1 mM ribose, $P > 0.9999$; no glucose/uridine and 1 mM ribose, $P = 0.3025$; no glucose/uridine and 10 mM ribose, ****$P < 0.0001$). ASPC1 (comparison between no glucose/uridine and no glucose + 1 mM uridine, ****$P < 0.0001$; no glucose/uridine and 1 mM glucose/no uridine, ****$P < 0.0001$; no glucose/uridine and 0.1 mM ribose, $P = 0.9974$; no glucose/uridine and 1 mM ribose, *$P = 0.0103$; no glucose/uridine and 10 mM ribose, ****$P < 0.0001$). The experiment was performed once. b,c, $n = 3$ biologically independent samples. Statistical significance was measured using two-tailed unpaired *t*-test. Intracellular: comparison between no uridine and 1 mM uridine: ***$P = 0.0005$ (uridine), ****$P < 0.0001$ (uracil); extracellular: comparison between no uridine and 1 mM uridine: ****$P < 0.0001$ (uridine), **$P = 0.008$ (uracil). d–f, $n = 3$ biologically independent samples per cell line. 'Others' indicates M other than M+0 or M+5, where applicable. Bars shown for PATU8988S are same as the WT bars (where applicable) for that cell line in the Extended Data Fig. 5. Tracing experiments were performed twice in these cells with similar results. g, Number of samples: sub-Q, tumours from 3 mice injected on the left and right flanks; ortho, tumours from 4 mice. Mode of uridine injection is intratumoural for sub-Q and intraperitoneal for ortho. h, Median concentration of uridine = 24.1 µM; median concentration of uracil = 90.2 µM; $n = 22$ biologically independent TIF samples. i, Median concentration of glucose = 3.71 mM (plasma) and 0.63 mM (TIF). $n = 8$ biologically independent plasma samples and 8 TIF samples extracted from 8 tumour samples from same mice. These samples are from the control group of the study in Fig. 4a. Statistical significance was measured with two-tailed unpaired *t*-test with Welch's correction, ****$P < 0.0001$. j,k, j shows the mass isotopologue distribution in uridine and k shows in the indicated metabolites. $n = 4$ biologically independent samples per group per cell line. 'Others' indicates M other than M+0 or M+5, where applicable. Data in a–k are shown as mean ± s.d. The metabolomics experiments (b–k) were performed once.

Figure 3. a, The experiment was performed once. b, $n = 4$ biologically independent samples per group per cell line. Statistical significance was measured using one-way ANOVA with Tukey's multiple comparisons test. PATU8988S, comparison between no uridine (−) and 1 mM uridine (+): ****$P < 0.0001$, $P = 0.9703$ and $P = 0.9089$ for WT, 1A and 1B groups, respectively. ASPC1, comparison between no uridine (−) and 1 mM uridine (+): ****$P < 0.0001$, $P > 0.9999$ and $P > 0.9999$ for WT, 1A and 1B groups, respectively. The experiments were performed three times with similar results. c, $n = 4$ biologically independent samples per group per cell line. Statistical significance was measured using one-way ANOVA with Tukey's multiple comparisons test. PATU8988S, comparison between no uridine (−) and 1 mM uridine (+): ****$P < 0.0001$, $P > 0.9999$ and $P = 0.9599$ for WT, 1A and 1B groups, respectively. ASPC1, comparison between no uridine (−) and 1 mM uridine (+): ****$P < 0.0001$, $P = 0.9977$ and $P = 0.6537$ for WT, 1A and 1B groups, respectively. The experiments were performed twice with similar results. d, $n = 3$ biologically independent samples per group. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. Comparison between WT and clonal cells 1A or 1B: ****$P < 0.0001$ (PATU8988S) and ***$P = 0.0003$ (ASPC1). Data are part of the metabolomics experiments shown in Extended Data Fig. 5a–c. The metabolomics experiment was performed once. e, $n = 3$ biologically independent samples per group. 'Others' indicates M other than M+0 or M+5, where applicable. Data are part of the metabolomics experiments shown in Extended Data Fig. 5e,h,j for ASPC1. The metabolomics experiment was performed once. f, Statistical significance was measured using two-tailed unpaired *t*-test with Welch's correction. Number of samples and statistical comparison: GSE62452 (NT, 61 vs PDA, 69, ***$P = 0001$), GSE71729 (middle: NT, 46 vs PDA, 145, *$P = 0.0466$), GSE71729 (right: primary, 145 vs liver met, PDA, 25, ****$P < 0.0001$). Box plot statistics: GSE42452 (NT: minimum = 3.582, maximum = 5.633, 25th percentile = 4.036, 75th percentile = 4.504, median = 4.262; PDA: minimum = 3.853, maximum = 5.989, 25th percentile = 4.37, 75th percentile = 4.843, median = 4.535); GSE71729 (NT: minimum = 2.18, maximum = 4.402, 25th percentile = 2.901, 75th percentile = 3.469, median = 3.139; PDA: minimum = 2.293, maximum = 4.725, 25th percentile = 3, 75th percentile = 3.657, median = 3.339); GSE71729 (primary: minimum = 2.293, maximum = 4.725, 25th percentile = 3, 75th percentile = 3.657,

median = 3.339; liver metastasis: minimum = 3.306, maximum = 5.768, 25th percentile = 3.564, 75th percentile = 4.498, median = 4.023). g, Representative images from patient 1 of 3 tumour tissues. PanCK, pan-cytokeratin, stain indicates tumour cells. i, Number of samples: *UPP1*-low, 144; *UPP1*-high, 144. j, Number of samples: no alteration, 43; G12D, 42. Statistical significance was measured using two-tailed unpaired *t*-test with Welch's correction, $^{**}P = 0.0029$. Box plot statistics: no alteration: minimum = 7.797, maximum = 10.66, median = 9.019, 25th percentile = 8.307, 75th percentile = 9.53; $KRAS^{G12D}$: minimum = 8.154, maximum = 11.3, median = 9.385, 25th percentile = 9.019, 75th percentile = 9.905. k, $n = 3$ biologically independent samples per cell line. Statistical significance was measured using two-tailed unpaired *t*-test. Comparison between Dox (−) and (+) in iKras* cell A9993: $^{***}P = 0.0002$; in iKras cell 8905: $^{**}P =$ 0088. The experiment was performed once. l, Vinculin is used as a loading control. The experiment was performed once. m, 3 biologically independent samples per group. Statistical significance was measured using two-tailed unpaired *t*-test. Comparison between cells cultured in uridine/glucose-containing medium with and without trametinib treatment: $^{****}P < 0.0001$; comparison between cells treated with and without trametinib in the presence of glucose but no uridine: $^{****}P < 0.0001$; comparison between cells treated with and without trametinib in the presence of uridine and no glucose: $^{****}P < 0.0001$; comparison between cells cultured with no uridine/glucose with and without trametinib treatment: $^{****}P < 0.0001$. The experiment was performed once. n, Vinculin is used as a loading control. The experiments were performed twice with similar results. o, Statistical significance was measured using one-way ANOVA with Tukey's multiple comparisons test. $n = 4$ biologically independent samples per group per cell line. PATU8988S (comparison between cells cultured with and without trametinib in the absence of uridine: $^{****}P < 0.0001$, and with uridine supplementation: $^{****}P < 0.0001$); DANG (comparison between cells cultured with and without trametinib in the absence of uridine: $P = 0.9967$, and with uridine supplementation: $^{****}P = 0.0001$); ASPC1 (comparison between cells cultured with and without trametinib in the absence of uridine: $P = 0.9987$, and with uridine supplementation: $^{***}P = 0.0001$. The experiment was performed once. Data in b–e,k,m,o are mean ± s.d.

Figure 4. a. The experiment involved a sequential treatment of mice with control IgG followed the next day by liposome PBS (control group) or anti-CSF1 followed the next day by clodronate (clod + Ab (macrophage depletion) group), after the establishment of palpable pancreatic orthotopic tumours with the KPC 7940b cell line. Data represent the average of quantification from three histological slides obtained per tumour. Sample size used for histology, $n = 5$ tumours from the same number of mice per group. Statistical significance was measured using two-tailed unpaired *t*-test with Welch's correction, $P = 0.0635$. The experiment was performed once. b. Sample size: control, 9; clod + Ab, 8 tumours from the corresponding number of mice. Statistical significance was measured using two-tailed unpaired *t*-test with Welch's correction, $^{**}P = 0.0046$. c, Plasma, $n = 8$; TIF, $n = 8$ control and macrophage-depleted ($n = 8$); tumours, $n = 8$ control and $n = 8$ macrophage-depleted. Statistical significance was measured using two-tailed unpaired *t*-test with Welch's correction. Comparison between control and clod + Ab (plasma uridine): $^{***}P = 0.0003$; control and clod + Ab (TIF uridine): $P = 0.7923$; control and clod + Ab (tumour uridine): $P = 0.1244$. d, The cells were cultured ±1 mM uridine in glucose-free medium supplemented with 2.5% dialysed FBS. Sample size = 4 biologically independent samples per group. Statistical significance was measured using two-tailed unpaired *t*-test. KPC 7940b, comparison of cell culture without and with 0.1 mM uridine: $^{****}P < 0.0001$, $P = 0.2577$ and $P = 0.1118$ for sgV, sg1 and sg3 groups, respectively; MT3-2D, comparison of cell culture without and with 0.1 mM uridine: $^{**}P = 0.0026$, $P = 0.9574$ and $P = 0.927$ for sgV, sg1 and sg3 groups, respectively. The experiments were performed twice with similar results. e, Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test, $n = 3$ biologically independent samples per group. Comparison between sgV and sg1 or sg3: $^{****}P < 0.0001$ (for both intracellular uridine and uracil). Extracellular uridine, comparison between sgV and sg1: $P = 0.1758$; comparison between sgV and sg3: $P = 0.1503$. Extracellular uracil, comparison between sgV and sg1 or sg3: $^{****}P < 0.0001$. g,h, $n = 3$ biologically independent samples per group. The metabolomics experiment was performed once. i, Tumour weight data are shown in j. j, Number of mice and tumour samples: sgV, $n = 6$; sg1, $n = 10$; sg3, $n = 8$. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. Comparison between sgV and sg1 or sg3: $^{****}P <$

0.0001. Experiment performed once. k, Number of samples: sgV, 8; sg1, 8; sg3, 8 tumours, corresponding to four mice per group. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. Comparison between sgV and sg1: $^{**}P = 0.003$; comparison between sgV and sg3: $^{***}P = 0.0002$. Experiment performed once. l, Samples used for metabolomics per group: sgV, 5; sg1, 6; sg3, 6. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. Uridine, comparison between sgV and sg1: $^{**}P = 0.0016$; sgV and sg3: $^{**}P = 0.0019$; uracil, comparison between sgV and sg1: $^{****}P < 0.0001$; sgV and sg3: $^{****}P < 0.0001$. Experiment performed once. Data in a–e,g,h,j–l are shown as mean ± s.d. Metabolites used for Venn diagrams (f,m,n) are significantly changed ($P < 0.05$) in the metabolomics profile of UPP1-KO compared with controls per cell line and were derived from LC–MS experiments (Fig. 3d and Extended Data Fig. 5a,c (PATU8988S and ASPC1, intracellular); Fig. 4e and Extended Data Fig. 10c (MT3-2D cell line, in vitro, intracellular); Fig. 4l, Extended Data Fig. 10g (MT3-2D tumours, in vivo) and Extended Data Fig. 10e,f (KPC 7940b tumours, in vivo). Statistical significance was determined using the limma package version 3.38.3 in R. The mouse schematic (a,i) was drawn with Adobe Illustrator 2021 version 25.4.3.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Human PDA data used in this study are publicly accessible under the accession numbers indicated in the 'PDA dataset analysis' and 'Pan-cancer data analysis' Methods sections, and GSE36133 (for CCLE data). The gene microarrays were obtained from NCBI Gene Expression Omnibus. The ICGC-AU microarray data (release_28) were downloaded from https://dcc.icgc.org/projects/ along with the associated clinical data and had no embargo (March 2020) and the TCGA data were downloaded from CBioPortal (https://www.cbioportal.org/). CCLE protein data were accessed

via DepMap portal (https://depmap.org/portal/). Human Protein Atlas data are available from https://www.proteinatlas.org/ENSG00000183696-UPP1/tissue/pancreas (normal) and https://www.proteinatlas.org/ENSG00000183696-UPP1/pathology/pancreatic+cancer#img (PDA). The accompanying source data are provided as supplementary tables. All other data that support the findings of this study are available from the corresponding authors upon request.

References

1. Singhi, A. D., Koay, E. J., Chari, S. T. & Maitra, A. Early detection of pancreatic cancer: opportunities and challenges. *Gastroenterology* **156**, 2024–2040 (2019).

2. Wood, L. D., Canto, M. I., Jaffee, E. M. & Simeone, D. M. Pancreatic cancer: pathogenesis, screening, diagnosis, and treatment. *Gastroenterology* **163**, 386–402.e1 (2022).

3. Ho, W. J., Jaffee, E. M. & Zheng, L. The tumour microenvironment in pancreatic cancer—clinical challenges and opportunities. *Nat. Rev. Clin. Oncol.* **17**, 527–540 (2020).

4. DuFort, C. C., DelGiorno, K. E. & Hingorani, S. R. Mounting pressure in the microenvironment: fluids, solids, and cells in pancreatic ductal adenocarcinoma. *Gastroenterology* **150**, 1545–1557.e2 (2016).

5. Encarnación-Rosado, J. & Kimmelman, A. C. Harnessing metabolic dependencies in pancreatic cancers. *Nat. Rev. Gastroenterol. Hepatol.* **18**, 482–492 (2021).

6. Halbrook, C. J. & Lyssiotis, C. A. Employing metabolism to improve the diagnosis and treatment of pancreatic cancer. *Cancer Cell* **31**, 5–19 (2017).

7. Helms, E., Onate, M. K. & Sherman, M. H. Fibroblast heterogeneity in the pancreatic tumor microenvironment. *Cancer Discov.* **10**, 648–656

(2020).

8. Koong, A. C. et al. Pancreatic tumors show high levels of hypoxia. *Int. J. Radiat. Oncol.* **48**, 919–922 (2000).

9. Beatty, G. L., Werba, G., Lyssiotis, C. A. & Simeone, D. M. The biological underpinnings of therapeutic resistance in pancreatic cancer. *Genes Dev.* **35**, 940–962 (2021).

10. Kamphorst, J. J. et al. Human pancreatic cancer tumors are nutrient poor and tumor cells actively scavenge extracellular protein. *Cancer Res.* **75**, 544–553 (2015).

11. Commisso, C. et al. Macropinocytosis of protein is an amino acid supply route in Ras-transformed cells. *Nature* **497**, 633–637 (2013).

12. Yang, S. et al. Pancreatic cancers require autophagy for tumor growth. *Genes Dev.* **25**, 717–729 (2011).

13. Zhao, H. et al. Tumor microenvironment derived exosomes pleiotropically modulate cancer cell metabolism. *eLife* **5**, e10250 (2016).

14. Halbrook, C. J. et al. Macrophage-released pyrimidines inhibit gemcitabine therapy in pancreatic cancer. *Cell Metab.* **29**, 1390–1399.e6 (2019).

15. Sousa, C. M. et al. Pancreatic stellate cells support tumour metabolism through autophagic alanine secretion. *Nature* **536**, 479–483 (2016).

16. Kim, P. K. et al. Hyaluronic acid fuels pancreatic cancer cell growth. *eLife* **10**, e62645 (2021).

17. Klijn, C. et al. A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312 (2015).

18. Possemato, R. et al. Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. *Nature* **476**, 346–350 (2011).

19. Barretina, J. et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

20. Choi, J. W. et al. Uridine protects cortical neurons from glucose deprivation-induced death: possible role of uridine phosphorylase. *J. Neurotrauma* **25**, 695–707 (2008).

21. Choi, J. W. et al. Uridine prevents the glucose deprivation-induced death of immunostimulated astrocytes via the action of uridine phosphorylase. *Neurosci. Res.* **56**, 111–118 (2006).

22. Wice, B. M., Reitzer, L. J. & Kennell, D. The continuous growth of vertebrate cells in the absence of sugar. *J. Biol. Chem.* **256**, 7812–7819 (1981).

23. Geiger, A. & Yamasaki, S. Cytidine and uridine requirement of the brain. *J. Neurochem.* **1**, 93–100 (1956).

24. Löffler, M., Wenzel, A. & Schneider, F. Cytokinetic studies on the switch from glucose to uridine metabolism, and vice versa, of Ehrlich ascites tumour cells in vitro. *Cell Prolif.* **20**, 181–190 (1987).

25. Linker, W., Löffler, M. & Schneider, F. Uridine, but not cytidine can sustain growth of Ehrlich ascites tumor cells in glucose-deprived medium with altered proliferation kinetics. *Eur. J. Cell Biol.* **36**, 176–181 (1985).

26. King, M. P. & Attardi, G. Human cells lacking mtDNA: repopulation with exogenous mitochondria by complementation. *Science* **246**, 500–503 (1989).

27. Son, J. et al. Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. *Nature* **496**, 101–105 (2013).

28. Lee, H.-J., Kremer, D. M., Sajjakulnukit, P., Zhang, L. & Lyssiotis, C. A. A large-scale analysis of targeted metabolomics data from

heterogeneous biological samples provides insights into metabolite dynamics. *Metabolomics* **15**, 103 (2019).

29. Yuan, M. et al. Ex vivo and in vivo stable isotope labelling of central carbon metabolism and related pathways with analysis by LC–MS/MS. *Nat. Protoc.* **14**, 313–330 (2019).

30. Sullivan, M. R. et al. Quantification of microenvironmental metabolites in murine cancers reveals determinants of tumor nutrient availability. *eLife* **8**, e44235 (2019).

31. Hezel, A. F., Kimmelman, A. C., Stanger, B. Z., Bardeesy, N. & DePinho, R. A. Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev.* **20**, 1218–1249 (2006).

32. Kerk, S. A., Papagiannakopoulos, T., Shah, Y. M. & Lyssiotis, C. A. Metabolic networks in mutant KRAS-driven tumours: tissue specificities and the microenvironment. *Nat. Rev. Cancer* **21**, 510–525 (2021).

33. Collins, M. A. et al. Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice. *J. Clin. Invest.* **122**, 639–653 (2012).

34. Ying, H. et al. Oncogenic Kras maintains pancreatic tumors through regulation of anabolic glucose metabolism. *Cell* **149**, 656–670 (2012).

35. Candido, J. B. et al. $CSF1R^+$ macrophages sustain pancreatic tumor growth through T cell suppression and maintenance of key gene programs that define the squamous subtype. *Cell Rep.* **23**, 1448–1460 (2018).

36. Zhang, Y. et al. Myeloid cells are required for PD-1/PD-L1 checkpoint activation and the establishment of an immunosuppressive environment in pancreatic cancer. *Gut* **66**, 124–136 (2017).

37. Zhang, D., Cao, D., Russell, R. & Pizzorno, G. p53-dependent suppression of uridine phosphorylase gene expression through direct

promoter interaction. *Cancer Res.* **61**, 6899–6905 (2001).

38. Dalin, S. et al. Deoxycytidine release from pancreatic stellate cells promotes gemcitabine resistance. *Cancer Res.* **79**, 5723–5733 (2019).

39. Tabata, S. et al. Thymidine catabolism as a metabolic strategy for cancer survival. *Cell Rep.* **19**, 1313–1321 (2017).

40. Wang, T. et al. Inosine is an alternative carbon source for $CD8^+$-T-cell function under glucose restriction. *Nat. Metab.* **2**, 635–647 (2020).

41. Jurkowitz, M. S., Litsky, M. L., Browning, M. J. & Hohl, C. M. Adenosine, inosine, and guanosine protect glial cells during glucose deprivation and mitochondrial inhibition: correlation between protection and ATP preservation. *J. Neurochem.* **71**, 535–548 (1998).

42. Litsky, M. L., Hohl, C. M., Lucas, J. H. & Jurkowitz, M. S. Inosine and guanosine preserve neuronal and glial cell viability in mouse spinal cord cultures during chemical hypoxia. *Brain Res.* **821**, 426–432 (1999).

43. Skinner. O. S. et al. Salvage of ribose from uridine or RNA supports glycolysis in nutrient-limiting conditions. *Nat. Metab.* https://doi.org/10.1038/s42255-023-00774-2 (2023).

44. Puleo, F. et al. Stratification of pancreatic ductal adenocarcinomas based on tumor and microenvironment features. *Gastroenterology* **155**, 1999–2013.e3 (2018).

45. Vaas, L. A. I. et al. opm: an R package for analysing OmniLog(R) phenotype microarray data. *Bioinformatics* **29**, 1823–1824 (2013).

46. Ran, F. A. et al. Genome engineering using the CRISPR–Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

47. Kerk, S. A. et al. Metabolic requirement for GOT2 in pancreatic cancer depends on environmental context. *eLife* **11**, e73245 (2022).

48. Yi, Z. et al. KDM6A regulates cell plasticity and pancreatic cancer progression by noncanonical activin pathway. *Cell. Mol. Gastroenterol. Hepatol.* **13**, 643–667 (2022).

49. Scales, M. K. et al. Combinatorial Gli activity directs immune infiltration and tumor growth in pancreatic cancer. *PLoS Genet.* **18**, e1010315 (2022).

50. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

51. Barrett, T. et al. NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* **41**, D991–D995 (2013).

52. Goldman, M. J. et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat. Biotechnol.* **38**, 675–678 (2020).

53. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576.e16 (2017).

54. Gearing, L. J. et al. CiiiDER: A tool for predicting and analysing transcription factor binding sites. *PLoS ONE* **14**, e0215495 (2019).

55. Steele, N. G. et al. Multimodal mapping of the tumor and peripheral blood immune landscape in human pancreatic cancer. *Nat. Cancer* **1**, 1097–1112 (2020).

Acknowledgements

We thank B. Bochner for providing the OmniLog instrument, helping with the data analysis, countless insightful discussions, and support; S. Chan for support with Biolog data collection and analysis; D. DeNardo and the DeNardo laboratory for support designing the macrophage-depletion studies; N. J. Guppy for processing and immunohistochemistry staining of mouse tumour samples; L. Howell for helping with the automated image analysis of immunohistochemistry staining; and members of the Sadanandam and Lyssiotis laboratories and the entire Pancreatic Disease

Initiative at the Rogel Cancer Center, University of Michigan, for their insightful comments and discussions. Z.C.N. was supported by the Michigan Postdoctoral Pioneer Program at the University of Michigan Medical School, NIH/NCI grant K99CA267176 and NIH/NIGMS grant R25GM143298. M.H.W. was supported by the CABRI Undergraduate Research Grant and NIH/NIGMS grant 5T32GM139774-02. P.S by NIH/NCI grants 5T32CA140044-12 and 5T32CA140044-13. M.R. was supported by NIH/HD grant T32-HD007505. C.J.H. was supported by NIH/NCI grant K99/R00CA241357. E.S.C. was supported by the Department of Veteran Affairs Career Development Award IK2BX005875. J.S. was supported by NIH/NCI grants K08CA234222 and R37CA262209. A.M. was supported by American Cancer Society (IRG-16-222-56), the University of Chicago Cancer Center Support Grant (P30 CA14599), the Pancreatic Cancer Action Network (2020 Career Development Award), the Brinson Foundation, the Cancer Research Foundation and the Ludwig Center for Metastasis Research. M.P. was supported by NIH/NCI grants R01CA151588 and R01CA198074. A.S. was supported by two grants from Pancreatic Cancer UK, including Future Leaders Fund (FLF2015_05_ICR) and Research Innovation Fund (RIF2014_06_Sadanandam) and funding from Ian Harty Charitable Trust. C.A.L. was supported by NIH/NCI grants R37CA237421, R01CA248160 and R01CA244931, and M.P.d.M. and C.A.L. were supported by UMCCC Core Grant P30CA046592. The funders had no role in study design, data collection and analysis, or the content and publication of this manuscript.

Author information

Author notes

1. These authors contributed equally: Zeribe C. Nwosu, Matthew H. Ward, Peter Sajjakulnukit, Pawan Poudel

Authors and Affiliations

1. Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI, USA

Zeribe C. Nwosu, Matthew H. Ward, Peter Sajjakulnukit, Steven Kasperek, Megan Radyk, Damien Sutton, Anthony Andren, Zachary Tolstyka, Ho-Joon Lee, Julia Ugras, Li Zhang, Christopher J. Halbrook & Costas A. Lyssiotis

2. Department of Chemistry, Washington University in St Louis, St Louis, MO, USA

Matthew H. Ward, Leah P. Shriver & Gary J. Patti

3. Department of Medicine, Washington University in St Louis, St Louis, MO, USA

Matthew H. Ward, Leah P. Shriver & Gary J. Patti

4. Center for Metabolomics and Isotope Tracing, Washington University in St Louis, St Louis, MO, USA

Matthew H. Ward, Leah P. Shriver & Gary J. Patti

5. Division of Molecular Pathology, The Institute of Cancer Research, London, UK

Pawan Poudel, Chanthirika Ragulan, Hari PS, Gift Nyamundanda & Anguraj Sadanandam

6. Cellular and Molecular Biology Program, University of Michigan, Ann Arbor, MI, USA

Rosa E. Menjivar

7. Ben May Department for Cancer Research, University of Chicago, Chicago, IL, USA

Juan J. Apiz-Saab, Lindsey N. Dzierozynski & Alexander Muir

8. Department of Surgery, University of Michigan, Ann Arbor, MI, USA

Kristee Brown, Xi He & Marina Pasca di Magliano

9. Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA

   Eileen S. Carpenter & Costas A. Lyssiotis

10. Department of Pathology and Clinical Labs, Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA

    Jiaqi Shi

11. Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA

    Marina Pasca di Magliano & Costas A. Lyssiotis

12. Centre for Global Oncology, Division of Molecular Pathology, The Institute of Cancer Research, London, UK

    Anguraj Sadanandam

Contributions

Z.C.N., M.H.W., A.S. and C.A.L. designed the study. Z.C.N., M.H.W. and C.A.L. wrote the manuscript. Z.C.N., M.H.W. and C.A.L. designed experiments, and Z.C.N., M.H.W. and P.S. collected data for the bulk of the experimental studies. Z.C.N. prepared manuscript figures. S.K. and Z.T. carried out the initial cell line screen, and P.P. conducted the analysis of the screening data that revealed lead metabolite–enzyme interactions. C.R., M.R., D.S., R.E.M., A.A., J.J.A.-S., K.B., H.-J.L., L.N.D., X.H., H.P., J.U., G.N., L.Z. and C.J.H. carried out experiments. A.A. and E.S.C. contributed to the data analysis. J.S., L.P.S., G.J.P., A.M., M.P.d.M., A.S. and C.A.L. provided resources, funding and conceptual input for experiments and supervised the research. All authors reviewed and approved the final manuscript.

Corresponding authors

Correspondence to Anguraj Sadanandam or Costas A. Lyssiotis.

Ethics declarations

Competing interests

C.A.L. has received consulting fees from Astellas Pharmaceuticals, Odyssey Therapeutics, and T-Knife Therapeutics, and is an inventor on patents pertaining to KRAS-regulated metabolic pathways, redox control pathways in pancreatic cancer, and targeting the GOT1-pathway as a therapeutic approach (US patent 2015126580-A1, 05/07/2015; US patent 20190136238, 05/09/2019; international patent WO2013177426-A2, 04/23/2015). A.S. received grants from Merck, Pierre Fabre and Bristol Myers Squibb and is an inventor on patents on colorectal cancer classification with differential prognosis and personalized therapeutic responses (PCT/IB2013/060416); prognostic and treatment response prediction in gastric cancer (priority patent CSC/BP7295892); patient classification and prognostic method (GEP-NET) (priority patent EP18425009.0); and molecular predictors of therapeutic response to specific anti-cancer agents (US9506926B2).

Peer review

Peer review information

Nature thanks the anonymous reviewers for their contribution to the peer review of this work. Peer review reports are available.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 Correlation of nutrient utilization with gene expression identifies uridine and UPP1.

a. Schematic overview of the parameters measured by the Biolog Phenotype Microarray. **b**. Heatmap showing the high confidence metabolites (HCMs), namely, the metabolites that were utilized above or below the median of negative controls as determined by one-tailed Wilcoxon rank sum test. Legend denotes fold change relative to median negative control signal, where red shows high utilization and blue shows low utilization. **c**. Spearman correlation plot, indicating the genes that showed positive or negative correlation with metabolites' RMA in the Biolog screen. **d**. Spearman correlations, r, between *UPP1* expression in cell lines (ref. [17]) and RMA of nucleosides that were included in the Biolog screen. n = 16 PDA cell lines. **e**. Heatmap showing the expression of glycolysis genes in human PDA tumours ranked based on *UPP1* expression (dataset: GSE71729, *UPP1* low, n = 72; high, n = 73). On the right: GSEA plot indicating the enrichment of glycolysis hallmark in the *UPP1* high relative to the low tumours. NES, normalized enrichment score. **f**. Downregulated pathways in PDA cell lines that metabolize uridine at a high level, as revealed by gene ontology (GO) analysis of the differentially expressed genes ($P < 0.05$). GO analysis was performed with DAVID (https://david.ncifcrf.gov/tools.jsp). Analysis was based on the differential genes derived from CCLE data and part of the data shown in Fig. 1g. **g**. GSEA plots of significantly enriched KEGG pathways in *UPP1*-high PDA tumours relative to *UPP1* low tumours. Plots are part of the data (e) from the analysis of GSE71729 human PDA dataset. **Statistics and reproducibility: a**, The kinetic measurement evaluated several parameters, including the time taken for cells to adapt to and catabolize a nutrient (lambda), the rate of uptake and catabolism (mu or slope), the total metabolic activity (area under the curve; AUC), and the maximum metabolic activity. The values from the maximum catabolic efficiency (maximum height, A) of the respective metabolites were used to determine relative metabolic activity (RMA).

Extended Data Fig. 2 Nutrient-deprived PDA use uridine to support metabolism.

a. Relative RMA upon uridine supplementation with or without glucose and glutamine. n = 4 biologically independent samples per group per cell line. **b**. Differential changed ($P < 0.05$) intracellular and **c**) extracellular metabolites

from PATU8988S cells supplemented with 1 mM uridine in glucose-free medium for 24 h, as determined by LC-MS metabolomics. n = 3 biologically independent samples per group. **d**. Differentially changed (*P* < 0.05) intracellular and **e**) extracellular metabolites from DANG cells supplemented with 1 mM uridine in glucose-free medium for 24 h, as determined by LC-MS metabolomics. n = 3 biologically independent samples per group. **f**. Intracellular and **g**) extracellular uridine and uracil from DANG cells supplemented with 1 mM uridine in glucose-free medium for 24 h, as determined by LC-MS. Plots in f-g are from the same experiment as d-e. n = 3 biologically independent samples per group. Statistical significance was measured using two-tailed unpaired t-test. Intracellular – comparison between no uridine and 1 mM uridine: *** P = 0.0001 for uridine, ** P = 0.0011 for uracil. Extracellular – comparison between no uridine and 1 mM uridine: *** P = 0.0002 for uridine, ** P = 0.0044 for uracil. **h**. Mass isotopologue distribution of $^{13}C_5$-uridine ribose-derived carbon in the displayed metabolites after 24 h culture in a glucose-free medium supplemented with 1 mM uridine. n = 3 biologically independent samples per cell line. Tracing experiments were performed twice in these cells with similar results. Data (a, f, g, h) are shown as mean ± s.d. ADP, adenosine diphosphate; AMP, adenosine monophosphate; GSSG, oxidized glutathione; NADH, nicotinamide adenine dinucleotide; UDP-GlcNAc, uridine diphosphate N-acetylglucosamine; X5P, xylulose 5-phosphate. **Statistics and Reproducibility: a**, n = 4 biologically independent samples per group per cell line. Statistical significance was measured using one-way ANOVA with Tukey's multiple comparisons test. PATU8988S (comparison between cells cultured in (−) glucose/glutamine/uridine and (−) glucose/glutamine/+uridine: *** P = 0.0007, comparison between (−) and (+) uridine in the presence of glutamine and without glucose: **** P < 0.00001, comparison between (−) and (+) uridine in the presence of glutamine and glucose: P = ns (0.8856)). DANG (comparison between cells cultured in (−) glucose/glutamine/uridine and (−) glucose/glutamine/+uridine: * P = 0.0165, comparison between (−) and (+) uridine in the presence of glutamine and without glucose: **** P < 0.0001, comparison between (−) and (+) uridine in the presence of glutamine and glucose: P = ns (0.7971)). ASPC1 (comparison between cells cultured in (−) glucose/glutamine/uridine and (−) glucose/glutamine/+uridine: **** P < 0.0001, comparison between (−) and

(+) uridine in the presence of glutamine and without glucose: **** $P < 0.00001$, comparison between (−) and (+) uridine in the presence of glutamine and glucose: P = ns (0.9968)).

Extended Data Fig. 3 PDA metabolize uridine via central carbon metabolism *in vitro* and *in vivo*.

a. Isotope tracing showing $^{13}C_5$-uridine ribose-derived carbon labelling in subcutaneous (Sub-Q) or orthotopically (Ortho) implanted KPC 7940b tumours collected 1 h after injecting the mice with 200 µL or 50 µL (Sub-Q) 0.2 M $^{13}C_5$-uridine. Number of samples: Sub-Q = 6 tumours from 3 mice injected on the left and right flanks; Ortho = 4 tumours from 4 mice. Mode of uridine injection is intratumoural for Sub-Q and intraperitoneal for Ortho. **b**. Mass isotopologue distribution of $^{13}C_5$-uridine ribose-derived carbon after 24 h culture of ASPC1 and PATU8988S cells in medium supplemented with 1 mM or 0.1 mM uridine, each with 5 mM or 0.1 mM glucose concentration. n = 4 biologically independent samples per group. M – mass; 'Others' – indicate M other than M+0 or M+5, where applicable. **c–d**. Isotope tracing showing metabolite labelling upon supplementation with $^{13}C_5$-uridine at the TIF uridine and glucose concentrations shown in Fig. 2h,i, after 12 h of culturing **c**) human PDA cell line ASPC1 and **d**) murine PDA cell line MT3-2D. The cell lines were cultured in medium supplemented with 25 µM $^{13}C_5$-uridine and 0.65 mM glucose. n = 3 biologically independent samples per cell line. AXP – AMP, ADP, ATP, and related metabolites; UXP – UMP, UDP, UTP and related metabolites. The experiments (a-d) were performed once. Data (a-d) are shown as mean ± s.d, where applicable. 2-PG, 2-phosphoglycerate; 6-PG, 6-phosphogluconate; ADP, adenosine diphosphate; AMP, adenosine monophosphate; ATP, adenosine triphosphate; DHAP, dihydroxyacetone phosphate; F1,6-BP, fructose 1,6-bisphosphate; CMP, cytidine monophosphate; G6P, glucose 6-phosphate; NAD+, nicotinamide adenine dinucleotide; R5P, ribose 5-phosphate; PEP, phosphoenolpyruvate; PPP, pentose phosphate pathway; TCA, tricarboxylic acid; S7P, sedoheptulose 7-phosphate; UDP, uridine diphosphate; UMP, uridine monophosphate; UTP,

uridine triphosphate; UDP-GlcNAc, uridine diphosphate N-acetylglucosamine.

Extended Data Fig. 4 Cellular pathways for ribose salvage from uridine.

a–c. Relative metabolic activity (RMA) of PDA cell lines depicting the preferential use of uridine at (**a,b**) low glucose concentration (0-1 mM) but not at **c**) the high glucose concentration (10 mM), over a 96 h culture. **d**. Schematic depicting metabolic pathways for uridine utilization. **e**. Expression of *PGM2*, *UCK1*, and *UCK2* in non-tumour (NT) and PDA tissue samples from the GSE71729 dataset. Number of samples: NT = 46, PDA = 145. **f–h**. Expression of *PGM2*, *UCK1* and *UCK2* in TCGA (human PDA tumour) and CCLE (human cell line) data separated into *UPP1* low (L) and high (H) subsets. **i**. Western blot for PGM2 in PDA cell lines. Presented in bold are cells that express high UPP1. These samples are the same batch as the data shown in Fig. 1e and the blot was generated during one of the technical replicates of that western blotting. kDa, unit for molecular weight. **j**. qPCR for *PGM2* in ASPC1 cells transfected with siPGM2 compared to non-targeting (siNT) control. On the right: RMA in *PGM2* knockdown cells +/− uridine (1 mM) or glucose (1 mM). **k**. qPCR for *UCK1* in ASPC1 cells transfected with siUCK1 compared to non-targeting (siNT) control. On the right: RMA in *UCK1* knockdown cells +/− uridine (1 mM) or glucose (1 mM). **l**. qPCR for *UCK2* in ASPC1 cells transfected with siUCK2 compared to non-targeting (siNT) control. On the right: RMA in *UCK2* knockdown cells +/− uridine (1 mM) or glucose (1 mM). **Statistics and reproducibility: a–c**, n = 3 biologically independent samples. Statistical significance for data in a-b was measured using one-way ANOVA with Tukey's multiple comparisons test. PATU8988S (comparison between no glucose and no glucose + uridine [0.1 mM]: **P = 0.0017; 0.01 mM glucose and 0.01 mM glucose + uridine: ***P = 0.0008; 0.1 mM glucose and 0.1 mM glucose + uridine: ***P = 0.0002; 1 mM glucose and 1 mM glucose + uridine: ****P < 0.0001). CAPAN2 (comparison between no glucose and no glucose + uridine: P = ns (0.5673); 0.01 mM glucose and 0.01 mM glucose + uridine: P = ns (0.0541); 0.1 mM glucose and 0.1 mM glucose + uridine: P = ns (0.092); 1 mM glucose and 1 mM glucose + uridine: ***P = 0.0007. #All four group comparisons have

significant P: 0.0468, 0.014, 0.0089, 0.0222, respectively, when directly compared using two-tailed unpaired t test). DANG (comparison between no glucose and no glucose + uridine: ****P < 0.0001; 0.01 mM glucose and 0.01 mM glucose + uridine: ****P < 0.0001; 0.1 mM glucose and 0.1 mM glucose + uridine: **P = 0.0051; 1 mM glucose and 1 mM glucose + uridine: **P = 0.0051). ASPC1 (comparison between no glucose and no glucose + uridine: *P = 0.0203; 0.01 mM glucose and 0.01 mM glucose + uridine: **P = 0.0031; 0.1 mM glucose and 0.1 mM glucose + uridine: ***P = 0.0003; 1 mM glucose and 1 mM glucose + uridine: ****P < 0.0001). Statistical significance for data in c was measured using two-tailed unpaired t test and P = ns (0.0852, 0.3509, 0.3021 and 0.3875 for PATU8988S, CAPAN2, DANG and ASPC1, respectively, in the comparison of no uridine vs 0.1 mM uridine groups in the presence of 10 mM glucose). **d**. Uridine can be channeled into DNA or RNA synthesis by direct phosphorylation from UCK1/2. Uridine can also be catabolized via UPP1 to produce uracil and ribose 1-phosphate. Ribose 1-phosphate is converted to ribose-5-phosphate by PGM2 and fuel pentose phosphate pathway, nucleotide biosynthesis and glycolysis. **e**. Statistical significance was measured using two-tailed unpaired t test with Welch's correction. Comparison between NT and PDA: *PGM2*, ****P < 0.0001; *UCK1*, ****P < 0.0001; *UCK2*, *P = 0.018. Box plot statistics – *PGM2* (NT: minima = 3.097, maxima = 5.527, median = 4.335, 25th percentile = 3.992, 75th percentile = 4.74; PDA: minima = 2.386, maxima = 6.433, 25th percentile = 4.424, median = 4.961, 75th percentile = 5.457), *UCK1* (NT: minima = 3.7, maxima = 5.1, median = 4.5, 25th percentile = 4.2, 75th percentile = 4.7; PDA: minima = 3.6, maxima = 5.1, median = 4.2, 25th percentile = 4, 75th percentile = 4.4), *UCK2* (NT: minima = 4.034, maxima = 7.615, median = 5.577, 25th percentile = 5.095, 75th percentile = 5.9; PDA: minima = 4.556, maxima = 6.93, median = 5.727, 25th percentile = 5.458, 75th percentile = 6.059). **f-h**. Number of samples: TCGA – *UPP1* low = 75, high = 75; CCLE – *UPP1* low = 22, high = 22. Statistical significance was measured using two-tailed unpaired t test with Welch's correction. Comparison between L and H groups in TCGA (*PGM2*: P = ns (0.1226), *UCK1*: P = ns (0.311); *UCK2*: *P = 0.0327). In the CCLE L versus H comparison, *PGM2*, *UCK1* and *UCK2* have P = ns (0.3486, 0.4645, 0.4381, respectively). TCGA – The Cancer Genome Atlas, CCLE – Cancer Cell Line Encyclopaedia. **i**. Vinculin is used as a loading control. **j**. Number of samples: qPCR = 3,

RMA = 3 biologically independent samples per group. qPCR – statistical significance was measured using unpaired t test; comparison between siNT and siPGM2: ***P = 0.0007. RMA – statistical significance was measured using multiple unpaired t tests with two-stage two-step method; comparison of siNT and siPGM2 in the presence of 1 mM glucose and no uridine: *P = 0.0452, and **P = 0.0014 in the presence of 1 mM uridine and no glucose. **k**. Number of samples: qPCR = 3, RMA = 3 biologically independent samples per group. qPCR – statistical significance was measured using two-tailed unpaired t test; comparison between siNT and siUCK1: ***P = 0.0004. RMA – statistical significance was measured using multiple unpaired t tests with two-stage two-step method. Comparison of siNT and siUCK1 knockdown samples in the presence of 1 mM glucose and no uridine: P = ns (0.8652), and P = ns (0.131) in the presence of 1 mM uridine and no glucose. **l**. Number of samples: qPCR = 3, RMA = 3 biologically independent samples per group. qPCR – statistical significance was measured using unpaired t test; comparison between siNT and siUCK2: ****P < 0.0001. RMA – statistical significance was measured using multiple unpaired t tests with two-stage two-step method; comparison of siNT and siUCK1 knockdown cells in the presence of 1 mM glucose and no uridine: *P = 0.035, and P = ns (0.8653) in the presence of 1 mM uridine and no glucose. Data (a–c, f–h, j–l) are shown as mean ± s.d. The experiments were performed once (a–c, k), and twice (j,l) with similar results.

Extended Data Fig. 5 UPP1 mediates the liberation of uridine-derived ribose for central carbon metabolism.

a. Relative intracellular uridine level in the UPP1 knockout PATU8988S cells (after 24 h) and ASPC1 cells (6 h) compared to the wild type supplemented with 1 mM uridine in medium with 10% dialyzed FBS and no glucose. n = 3 biologically independent samples per group per cell line. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. PATU8988S (intracellular uridine) – comparison between WT and 1A, ***P = 0.0002; comparison between WT and 1B, ****P < 0.0001. ASPC1 (intracellular uridine) – comparison between WT and 1A or 1B, ****P < 0.0001. **b**. Relative extracellular uridine and uracil in UPP1 knockout ASPC1 cells compared to the wild

type supplemented with 1 mM uridine for 6 h in medium with 10% dialyzed FBS and no glucose. n = 3 biologically independent samples per group. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. Extracellular uridine: comparison between WT and 1A or 1B, ****P < 0.0001; extracellular uracil: comparison between WT and 1A or 1B, ****P < 0.0001. **c**. Heatmaps of significantly altered intracellular metabolites (P < 0.05) in PATU8988S cells after 24 h (left) and ASPC1 after 6 h (right), as measured by LC-MS. Data used for the intracellular uridine/uracil plots (a-b) were extracted from this profiling study. **d–k**. Mass isotopologue distribution of 1 mM $^{13}C_5$-uridine ribose-derived carbon into the indicated metabolic pathways in wildtype (WT) or UPP1-KO PATU8988S and ASPC1 cells. M – mass; 'Others' – indicate M other than M+0 or M+5, where applicable. 1A and 1B denote UPP1-KO sgRNA clones. Data (a–b, d–k) are shown as mean ± s.d. Metabolomics experiments were done once. **Statistics and reproducibility:** Abbreviations – ADP, adenosine diphosphate; ATP, adenosine triphosphate; AMP, adenosine monophosphate; GSSG, oxidized glutathione; NAD+ nicotinamide adenine dinucleotide; PEP, phosphoenolpyruvate; TCA cycle, tricarboxylic acid cycle; UDP-GlcNAc, uridine diphosphate N-acetylglucosamine; UMP, uridine monophosphate; UTP, uridine triphosphate.

Extended Data Fig. 6 UPP1 expression is elevated in PDA and other cancer types.

a. TCGA RNA seq data showing the expression of *UPP1* and its paralog *UPP2*. FPKM, fragments per kilobase of exon per million mapped fragments. **b**. RNA seq showing UPP1 expression in various normal human tissues (Human Protein Atlas data), as obtained from the National Center for Biotechnology Information (NCBI) portal. **c**. Histological data showing UPP1 protein expression in normal pancreatic tissue compared to PDA. **d**. *UPP1* expression in human non-PDA cancers accessed in publicly accessible datasets. **e**. Kaplan-Meier plot of survival probability (log-rank test) as obtained from KM-plotter (https://kmplot.com/analysis/) using the default parameters. **f**. Relative Metabolic Activity (RMA), reflecting NADH levels, in human non-PDA cancer cell lines supplemented with uridine (as indicated) in 1 mM glucose medium. n = 5 biologically

independent samples per group per cell line. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. A549 (lung cancer cell line, comparison between no uridine and 0.1 mM uridine, *P = 0.0405 or 1 mM uridine, ***P = 0.0004); HT1080 (fibrosarcoma cell line, comparison between no uridine and 0.1 mM uridine, P = ns (0.1773) or 1 mM uridine, ***P = 0.0001); HCT116 (colon cancer cell line, comparison between no uridine and 0.1 mM uridine, *P = 0.0294 or 1 mM uridine, **P = 0.0081); U2OS (osteosarcoma cell line, comparison between no uridine and 0.1 mM uridine, **P = 0.002 or 1 mM uridine, ****P < 0.0001). **Statistics and reproducibility: a**, n = 150 samples each. **b**. Sample size, n: salivary gland = 3, pancreas = 2, ovary = 2, skin = 3, prostate = 4, stomach = 3, kidney = 4, testis = 7, small intestine = 4, fat = 3, endometrium = 3, thyroid = 4, liver = 3, urinary bladder = 2, lymph node = 5, brain = 3, duodenum = 2, colon = 5, placenta = 4, heart = 4, spleen = 4, gall bladder = 3, lung = 5, adrenal = 3, appendix = 3, esophagus = 3, bone marrow = 4. RPKM, reads per kilobase of exon per million reads mapped. *UPP1* expression in normal pancreas is extremely low (second lowest of the > 25 tissues compared). **c**. Data obtained from the Human Protein Atlas (URL for 'Normal' - https://www.proteinatlas.org/ENSG00000183696-UPP1/tissue/pancreas; PDA – https://www.proteinatlas.org/ENSG00000183696-UPP1/pathology/pancreatic+cancer#img). **d**. Sample size, n: NT = 19, tumour = 408 (bladder cancer, TCGA); NT = 5, tumour = 154 (glioblastoma, TCGA); NT = 44, tumour = 520 (head and neck cancer, TCGA); NT = 59, tumour = 551 (lung cancer, TCGA); NT = 11, tumour = 184 (oesophageal cancer, TCGA); NT = 52, tumour = 497 (prostate cancer, TCGA); NT = 41, tumour = 452 (colon cancer); health colon mucosa = 50, distant colon = 98, tumour = 98 (colon cancer, GSE44076). NT – non-tumour/adjacent normal tissue. Data (a-b, f) shown as mean ± s.d. The experiments were performed three times with similar results. Box plot statistics – TCGA, bladder carcinoma (primary: minima = 5.83, maxima = 13.5, median = 9.77, 25th percentile = 9.015, 75th percentile = 10.47; normal: minima = 6.61, maxima = 12.43, median = 8.35, 26th percentile = 8.03, 75th percentile = 9.59); glioblastoma multiforme (primary: minima = 5.71, maxima = 11.84, median = 9.585, 25th percentile = 8.79, 75th percentile = 10.143; normal: minima = 7.04, maxima = 7.63, median = 7.4, 25th percentile = 7.36, 75th percentile = 7.61); head and neck squamous cell carcinoma (primary: minima = 6.59,

maxima = 15.64, median = 10.75, 25th percentile = 9.787, 75th percentile = 11.565; normal: minima = 6.38, maxima = 13.73, median = 10.42, 25th percentile = 8.672, 75th percentile = 11.065); lung adenocarcinoma (primary: minima = 6.45, maxima = 13.44, median = 9.8, 25th percentile = 9.13, 75th percentile = 10.49; normal: minima = 8.3, maxima = 11.39, median = 9.3, 25th percentile = 8.945, 75th percentile = 9.93); esophageal carcinoma (primary: minima = 6.7, maxima = 13.08, median = 9.26, 25th percentile = 8.578, 75th percentile = 10.21; normal: minima = 6.17, maxima = 12.39, median = 7.62, 25th percentile = 6.7, 75th percentile = 8.26); prostate adenocarcinoma (primary: minima = 3.96, maxima = 9.69, median = 6.58, 25th percentile = 5.98, 75th percentile = 7.14; normal: minima = 4.56, maxima = 8.62, median = 6.97, 25th percentile = 6.447, 75th percentile = 7.24); colon cancer (primary: minima = 6.41, maxima = 12.96, median = 8.535, 25th percentile = 8.068, 75th percentile = 9.07; normal: minima = 7.76, maxima = 11.29, median = 9.57, 25th percentile = 9.09, 75th percentile = 9.92). Colon cancer (GSE44076, primary: minima = 4.564, maxima = 7.608, median = 5.917, 25th percentile = 5.487, 75th percentile = 6.405; normal: minima = 4.568, maxima = 9.154, median = 7.18, 25th percentile = 6.781, 75th percentile = 7.824; healthy colon mucosal cells: minima = 5.884, maxima = 8.279, median = 7.529, 25th percentile = 7.153, 75th percentile = 7.74). Statistical significance was tested using two-sided Wilcoxon or Kruskal-Wallis tests.

Extended Data Fig. 7 UPP1 is expressed in PDA and TME cells and predicts survival outcome.

a. RNAscope images showing UPP1 expression in tumour (PDA) compared to the adjacent non-tumour tissues. Pan-cytokeratin (PanCK) indicates the tumour cells; DAPI, nuclear stain. The images are representative of three 20x acquisitions per tissue slide, and of two independent experiments. Scale bar indicates 100 µm. **b**. UMAP plot showing the expression of *UPP1* at the transcript level, as determined by single cell RNA seq of PDA tissues from two patients (#1238 and 1302). **c**. Violin plots showing *UPP1* expression in various tumour microenvironment cell types, including myeloid and epithelial cells where *UPP1* is highest. Right, UMAP plot showing the specific cell compartments expressing *UPP1* for all patients' samples combined (n = 16). Data used in plots b-c are from a previously

published dataset[55]. **d**. Immunohistochemistry of UPP1 in patient biopsy sections from previously published tissue microarray[48]. Micrographs are representative from 213 patient samples in the microarray and two independent experiments. Large scale bar indicates 100 µm; scale bar on insets indicates 25 µm. **e**. Kaplan-Meier plot showing survival probability (log-rank test) based on *UPP1* expression in three separate datasets. Each dataset was split into two – *UPP1* high and *UPP1* low – based on the ranked *UPP1* expression value. Sample size: low = 133, high = 134 (ICGC); low = 62, high = 63 (GSE71729), low = 73, high = 73 (TCGA). TME – tumour microenvironment.

Extended Data Fig. 8 KRAS-MAPK pathway activation and nutrient availability drive UPP1 expression.

a. Normalized UPP1 protein expression in Kras wildtype and mutant cell lines based on CCLE protein data accessed via the DepMap portal. **b**. *Upp1* mRNA expression in iKras* orthotopic tumours and cell lines from dataset GSE32277. **c-d**. Western blot showing UPP1 expression in human PDA cell lines **c**) ASPC1 cells and **d**) DANG and PATU8988S after 24 h culture +/– trametinib [MEK inhibitor], uridine, or glucose. kDa, unit for molecular weight. **e**. qPCR for *UPP1* in human PDA cell lines DANG and PATU8988S treated for 24 h with trametinib (1 µM), uridine (1 mM), or 5 mM glucose. **f**. Western blot showing pERK upon treatment of murine iKras 9805 cells for 24 h with trametinib or doxycycline (1 µg/mL). **g**. qPCR for *Upp1* in iKras* 9805 mouse PDA cells, 24 h after treatment with trametinib or doxycycline (1 µg/mL). **h**. qPCR for *UPP1* in ASPC1 cells treated for 48 h with trametinib and at 0.1 mM uridine or 1 mM glucose concentrations. **i**. Western blot for UPP1 and pERK treated for 48 h with trametinib and low glucose [1 mM] and near physiological uridine concentration [0.1 mM]. **j**. Densitometric quantification of pERK and UPP1 in the ASPC1 blots shown in Fig. 3n. **k**. Metabolomics profiling showing the spectrum of metabolic changes induced in ASPC1 upon pERK inhibition with trametinib [1 µM], 24 h after culture. **l**. MTT assay showing relative proliferation of PDA cell lines treated with 1.25 µM trametinib [MEK inhibitor] +/− 1 mM uridine in the presence of glucose [5 mM] at 96 h. **Statistics and reproducibility: a**, Sample size – wild type 0 and mutation 1: 304 and 69 (pan-cancer), 15 and 15 (colon cancer), 54 and 25

(lung cancer). Box plot statistics – pan-cancer (KRAS = 0: minima = −4.8237, maxima = 4.8004, median = −0.3029, 25th percentile = −1.3212, 75th percentile = 0.7687; minima = −2.1805, maxima = 3.7445, median = −0.3019, 25th percentile = −0.815, 75th percentile = 1.2867), colon (KRAS = 0: minima = −2.3253, maxima = 3.1201, median = -0.5127, 25th percentile = −0.8908, 75th percentile = −0.0958; KRAS = 1: minima = −2.1805, maxima = 0.2672, median = −0.9153, 25th percentile = −1.3926, 75th percentile = −0.5595), lung (KRAS = 0: minima = −2.7979, maxima = 4.0884, median = 0.4201, 25th percentile = -1.305, 75th percentile = 0.6032; KRAS = 1: minima = −0.9357, maxima = 3.7445, median = 0.8446, 25th percentile = −0.4923, 75th percentile = 1.3752). **b**. Sample size: tumours (Kras_OFF = 9, Kras_ON = 10), cell lines (Kras_OFF = 5, Kras_ON = 5). Statistical significance was measured using two-tailed unpaired t test. Comparison of Kras_OFF to Kras_ON: **P = 0.0088 (tumours) and *P = 0.0244 (cell lines). **c-d**. pERK is used as a readout for MAPK pathway induction/activity. ERK and Vinculin are used as loading controls. Blots are representative of two biological and technical replicates for ASPC1 and one biological replicate for PATU8988S and DANG with similar results. **e**. n = 3 biologically independent samples per group. Statistical significance was measured with one-way ANOVA with Tukey's multiple comparisons test. Comparisons between groups for DANG (from left to right): P = ns (0.9097), P = ns (0.5014), **P = 0.0025 and ****P < 0.0001. Comparisons between groups for PATU8988S (from left to right): ***P = 0.0002, ***P = 0.001, ****P < 0.0001 and ****P < 0.0001. **f**. Vinculin is used as a loading control. **g**. n = 3 biologically independent samples per group. Statistical significance was measured with one-way ANOVA with Dunnett's multiple comparisons test. Comparison between (–) and (+) doxycycline, ****P < 0.0001. Comparison between (–) doxycycline: vs 10 μM trametinib, P = ns (0.9997); 10 μM trametinib + doxycycline, P = ns (0.8226); 1 μM trametinib, P = ns (0.9997); 1 μM trametinib + doxycycline, P = ns (0.9994); 0.1 μM trametinib, P = ns (0.1904); 0.1 μM trametinib + doxycycline, P = ns (>0.9999). **h**. n = 3 biologically independent samples per group. Statistical significance was measured with one-way ANOVA with Tukey's multiple comparisons test. Comparisons between groups (from left to right): ****P < 0.0001, ****P < 0.0001, ****P < 0.0001 and ****P < 0.0001. The experiments (e, g, h) were performed once with similar results on UPP1 displayed by the three cell lines. **i**. n = 3

biologically independent samples per group. This blot was run on the same gel as Fig. 3n hence the first two columns (separated by a box) overlap between the two blots. **j**. Blots (c,i) are representative of two independent experiments; blot e experiment was done once. **k**. n = 3 biologically independent samples per group. The statistical significance (P < 0.05) was determined using *limma* package version 3.38.3 in R. **l**. Statistical significance was measured using one-way ANOVA with Tukey's multiple comparisons test. n = 4 biologically independent samples per group. PATU8988S (comparison between cells cultured with and without trametinib in the absence of uridine: ****P < 0.0001, and with uridine supplementation: ****P < 0.0001); DANG (comparison between cells cultured with and without trametinib in the absence of uridine: **P = 0.0055, and with uridine supplementation: ***P = 0.0009); ASPC1 (comparison between cells cultured with and without trametinib in the absence of uridine: P = not significant, and with uridine supplementation: ****P = 0.0006). Data (b, e, g-h, l) shown as mean ± s.d.

Extended Data Fig. 9 Regulation of *UPP1* expression is independent of c-MYC.

a. Western blot showing Myc inhibition by 10058-F4 in ASPC1 cells after 24 h of culture. On the right: *UPP1* mRNA expression determined by qPCR. kDa, unit for molecular weight. **b**. Western blot of Myc inhibition by Fedratinib in ASPC1 cells after 24 h of culture. On the right: *UPP1* mRNA expression determined by qPCR. **c**. CiiDER analysis of transcription factor binding sites in the promoters of mouse and human UPP1. Myc binding sites were not detected. Details of analysis is in the "Methods" section. **d**. qPCR showing *UPP1* expression upon uridine supplementation with or without basal glucose concentration in culture medium. **e**. RNA seq data showing the expression of *Upp1* in sorted tumour cells and in KPC cells cultured in vitro in regular RPMI culture medium or tumour interstitial fluid medium (TIFM). Sample sizes: Tumour, n = 6, RPMI, n = 3 biologically independent cell samples, TIFM, n = 3 biologically independent cell samples. Normalized by log transformation [log2 (count +1)]. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. Comparison between RPMI and TIFM group, ****P < 0.0001 or tumour group, ****P < 0.0001. **Statistics and reproducibility: a**,

qPCR, n = 3 biologically independent samples per group. **b**. n = 3 biologically independent samples per group. Blots shown (a-b) are representative of two biological and technical replicate analyses with similar results. **d**. n = 3 biologically independent samples per group per cell line. Statistical significance was measured using one-way ANOVA with Tukey's multiple comparisons test. PATU8988S (comparison between no glucose and no glucose + 1 mM uridine: P = ns (0.994); comparison between cells cultured in glucose-containing medium with and without uridine: **P = 0.005); comparison between no glucose and glucose: *P = 0.0316; CAPAN2 (comparison between no glucose and no glucose + 1 mM uridine: ****P < 0.0001; comparison between cells cultured in glucose-containing medium with and without uridine: P = ns (0.8688); DANG (comparison between no glucose and no glucose + 1 mM uridine: ****P < 0.0001; comparison between cells cultured in glucose-containing medium with and without uridine: *P = 0.0021); ASPC1 (comparison between no glucose and no glucose + 1 mM uridine: ****P < 0.0001; comparison between cells cultured in glucose-containing medium with and without uridine: P = ns (0.5339). Comparison between no glucose and glucose for CAPAN2, DANG and ASPC1: ****P < 0.0001. Data (a,b,d,e) shown as mean ± s.d.

Extended Data Fig. 10 Knockout of UPP1 suppresses *in vivo* uridine catabolism and tumour growth.

a. LC-MS analysis of extracellular uridine and uracil in unpolarized bone marrow-derived macrophages (M0; 10 ng/mL M-CSF), those polarized toward an M1 fate (10 ng/mL LPS), an M2 fate (10 ng/mL murine IL-4), or a tumour-educated (TEM) phenotype with 75% PDA conditioned medium and compared to growth medium (DMEM). n = 3 biologically independent samples per group. Data was extracted from a previously published metabolomics[14]. **b**. Relative uracil abundance in plasma, tumour interstitial fluid (TIF), and bulk tumour from the experiment described in Fig. 4a, where Clod + Ab indicates the macrophage-depleted group. **c**. Extracellular (left) and intracellular (right) profiles of significantly changed (P < 0.05) metabolites from sgV and UPP1-KO (sg1, sg3) MT3-2D cells. The cells were cultured in medium with 1 mM uridine and no glucose for 24 h, as determined by metabolomics. **d**. Proliferation assay of sgVector (sgV) and

UPP1-KO (sg1, sg3) MT3-2D and KPC 7940b mouse PDA cell lines cultured over 72 h and 96 h, respectively, in normal growth medium with 10% dialyzed FBS. **e**. Tumour weight and photograph and bulk tumour uridine and uracil from the orthotopic implantation of sgV and UPP1-KO (sg3) KPC 7940b cell lines into C57BL/6J mice; see Fig. 4i for experimental details. **f,g**. Tumour metabolomics profile indicating statistically significant ($P < 0.05$) metabolites from sgV and UPP1-KO **f**) KPC 7940b and **g**) MT3-2D-derived tumours. **h**. Immunohistochemistry (IHC) staining of macrophages (F4/80), large blood vessels (CD31+), and cytotoxic cells (CD8) using tissue sections of MT3-2D orthotopic tumours. Micrographs are representative of 10 fields per image obtained per experiment group. Scale bar 500 µm. On the right is the respective quantification of each IHC stain. **Statistics and reproducibility: a**, Statistical significance was measured using one-way ANOVA and Dunnett's multiple comparisons test. Uridine – DMEM: comparison to M0 Mac group ****P < 0.0001; to M1 Mac group P = ns (0.0758); to M2 Mac group ****P < 0.0001; to TEM group ****P < 0.0001. Uracil – DMEM: comparison to M0 Mac P=ns (0.5476); to M1 Mac group ****P < 0.0001; to M2 Mac group P=ns (0.9894); to TEM group ****P < 0.0001. **b**. Statistical significance was measured using a two-tailed unpaired t-test – Plasma uracil **P = 0.0045 with n = 8 biologically independent replicates in both groups, TIF uracil P = ns (0.6872) with n = 8 biologically independent replicates in both groups– and using Welch's t-test P=ns (0.9682) with n = 8 biologically independent replicates in both groups. **c**. n = 3 biologically independent samples per group. Color scale denotes fold change. Below: Venn diagram showing overlapping metabolites that accumulated in both human (PATU8988S and ASPC1) and mouse (MT3-2D) cell lines upon UPP1 knockout. **d**. MT3-2D, n = 4 and KPC 7940b, n = 3 biologically independent samples per group. **e**. n = 8 biologically independent samples per group. Statistical significance measured using two-tailed unpaired t-test with Welch's correction, **P = 0.0059. On the right: bulk tumour uridine and uracil as measured using metabolomics. Sample size: sgV = 8, sg3 = 7. Uridine: ****P < 0.0001. Uracil: ****P < 0.0001. **f,g**. Sample size: KPC 7940b sgV = 8 and sg3 = 7, MT3-2D sgV = 5 sg1 = 6 and sg3 = 6 biologically independent samples from mice with orthotopic pancreatic tumours. **h**. Sample size: F4/80 sgV = 7, sg1 = 6, sg3 = 5; CD31 n = 6 per group; CD8+ sgV = 6, sg1 = 5, sg3 = 6, all biologically independent samples

per group. Statistical significance was measured using one-way ANOVA with Dunnett's multiple comparisons test. Comparison between % F4/80 sgV and sg1 P = ns (>0.9968), and sgV and sg3 P = ns (0.9583). Data (a, b, d, e) shown as mean ± s.d; horizontal bars in h represent mean value.

Supplementary information

Reporting Summary

Peer Review File

Supplementary Figure 1

Full, uncropped western images for Fig. 1e.

Supplementary Figure 2

Full, uncropped western images for Fig. 3a.

Supplementary Figure 3

Full, uncropped western images for Fig. 3l.

Supplementary Figure 4

Full, uncropped western images for Fig. 3n.

Supplementary Figure 5

Full, uncropped western images for Extended Data Fig. 4i.

Supplementary Figure 6

Full, uncropped western images for Extended Data Fig. 8c.

Supplementary Figure 7

Full, uncropped western images for DANG cell line in Extended Data Fig. 8d.

Supplementary Figure 8

Full, uncropped western image for TU8988S cell line in Extended Data Fig. 8d.

Supplementary Figure 9

Full, uncropped western images for Extended Data Fig. 8f.

Supplementary Figure 10

Full, uncropped western images for Extended Data Fig. 8i.

Supplementary Figure 11

Full, uncropped western images for Extended Data Fig. 9a,b.

Supplementary Table 1

The nutrient panel included in the Biolog screen.

Supplementary Table 2

CCLE expression profile of PDA cells identified as uridine high versus low consumers/metabolizers.

Supplementary Table 3

Pathway annotation of differentially expressed genes in uridine-high consumers/metabolizers.

Supplementary Table 4

sgRNA sequences used in this study.

Supplementary Table 5

qPCR primers used in this study.

Supplementary Table 6

Source data for metabolomics, mouse studies and qPCR. Full source data for all the metabolomics and tumour studies are presented, labelled by sub-figure and separated by tabs. Additionally, mRNA expression is provided for *PGM2*, *UCK1* and *UCK2* following siRNA-mediated knockdown.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Nwosu, Z.C., Ward, M.H., Sajjakulnukit, P. *et al.* Uridine-derived ribose fuels glucose-restricted pancreatic cancer. *Nature* **618**, 151–158 (2023). https://doi.org/10.1038/s41586-023-06073-w

- Received: 21 June 2021
- Accepted: 12 April 2023
- Published: 17 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06073-w

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

This article is cited by

- **Salvage of ribose from uridine or RNA supports glycolysis in nutrient-limited conditions**
    - Owen S. Skinner
    - Joan Blanco-Fernández
    - Alexis A. Jourdain

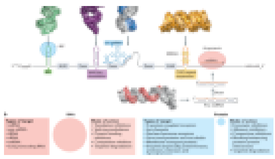
  Nature Metabolism (2023)

Salvage of ribose from uridine or RNA supports glycolysis in nutrient-limited conditions

- Owen S. Skinner
- Joan Blanco-Fernández
- Alexis A. Jourdain

Nature Metabolism Letter Open Access 17 May 2023

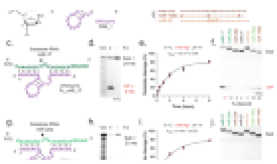
This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06073-w



Targeting RNA structures with small molecules

08 August 2022

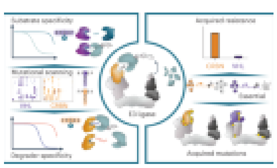
Jessica L. Childs-Disney, Xueyi Yang, … Matthew D. Disney



Targeting non-coding RNA family members with artificial endonuclease XNAzymes

24 September 2022

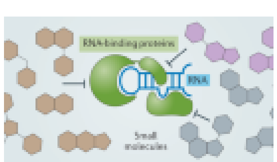
Maria J. Donde, Adam M. Rochussen, … Alexander I. Taylor



Functional E3 ligase hotspots and resistance mechanisms to small-molecule degraders

03 November 2022

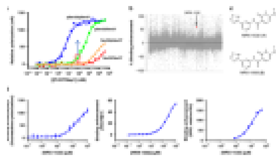
Alexander Hanzl, Ryan Casement, … Georg E. Winter



Inhibition of RNA-binding proteins with small molecules

15 July 2020

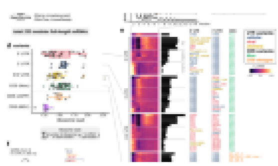
Peng Wu



Prospective discovery of small molecule enhancers of an E3 ligase-substrate interaction

29 March 2019

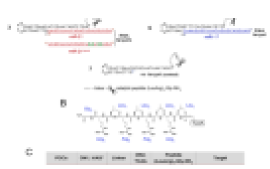
Kyle R. Simonetta, Joshua Taygerly, … Anjanabha Saha



Combinatorial optimization of mRNA structure, stability, and translation for RNA-based therapeutics

22 March 2022

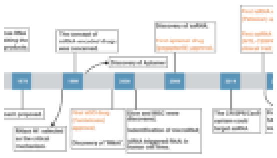
Kathrin Leppek, Gun Woo Byeon, … Rhiju Das



Peptide-oligonucleotide conjugates exhibiting pyrimidine-X cleavage specificity efficiently silence miRNA target acting synergistically with RNase H

09 October 2018

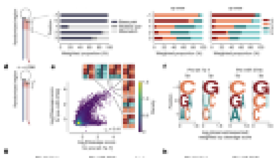
O. A. Patutina, M. A. Bazhenov, … M. A. Zenkova



RNA-based therapeutics: an overview and prospectus

23 July 2022

Yiran Zhu, Liyuan Zhu, … Hongchuan Jin



Sequence determinant of small RNA production by DICER

22 February 2023

Young-Yoon Lee, Haedong Kim & V. Narry Kim

- Article
- Open Access
- Published: 24 May 2023

Programming inactive RNA-binding small molecules into bioactive degraders

- Yuquan Tong,
- Yeongju Lee,

- Xiaohui Liu,
- Jessica L. Childs-Disney,
- Blessy M. Suresh,
- Raphael I. Benhamou,
- Chunying Yang,
- Weimin Li,
- Matthew G. Costales,
- Hafeez S. Haniff,
- Sonja Sievers,
- Daniel Abegg,
- Tristan Wegner,
- Tiffany O. Paulisch,
- Elizabeth Lekah,
- Maison Grefe,
- Gogce Crynen,
- Montina Van Meter,
- Tenghui Wang,
- Quentin M. R. Gibaut,
- John L. Cleveland,
- Alexander Adibekian,
- Frank Glorius ,
- Herbert Waldmann &
- …
- Matthew D. Disney

Show authors

Nature **volume 618**, pages 169–179 (2023)

- 18k Accesses
- 230 Altmetric
- Metrics details

Abstract

Target occupancy is often insufficient to elicit biological activity, particularly for RNA, compounded by the longstanding challenges surrounding the molecular recognition of RNA structures by small molecules. Here we studied molecular recognition patterns between a natural-product-inspired small-molecule collection and three-dimensionally folded RNA structures. Mapping these interaction landscapes across the human transcriptome defined structure–activity relationships. Although RNA-binding compounds that bind to functional sites were expected to elicit a biological response, most identified interactions were predicted to be biologically inert as they bind elsewhere. We reasoned that, for such cases, an alternative strategy to modulate RNA biology is to cleave the target through a ribonuclease-targeting chimera, where an RNA-binding molecule is appended to a heterocycle that binds to and locally activates RNase L[1]. Overlay of the substrate specificity for RNase L with the binding landscape of small molecules revealed many favourable candidate binders that might be bioactive when converted into degraders. We provide a proof of concept, designing selective degraders for the precursor to the disease-associated microRNA-155 (pre-miR-155), *JUN* mRNA and *MYC* mRNA. Thus, small-molecule RNA-targeted degradation can be leveraged to convert strong, yet inactive, binding interactions into potent and specific modulators of RNA function.

Main

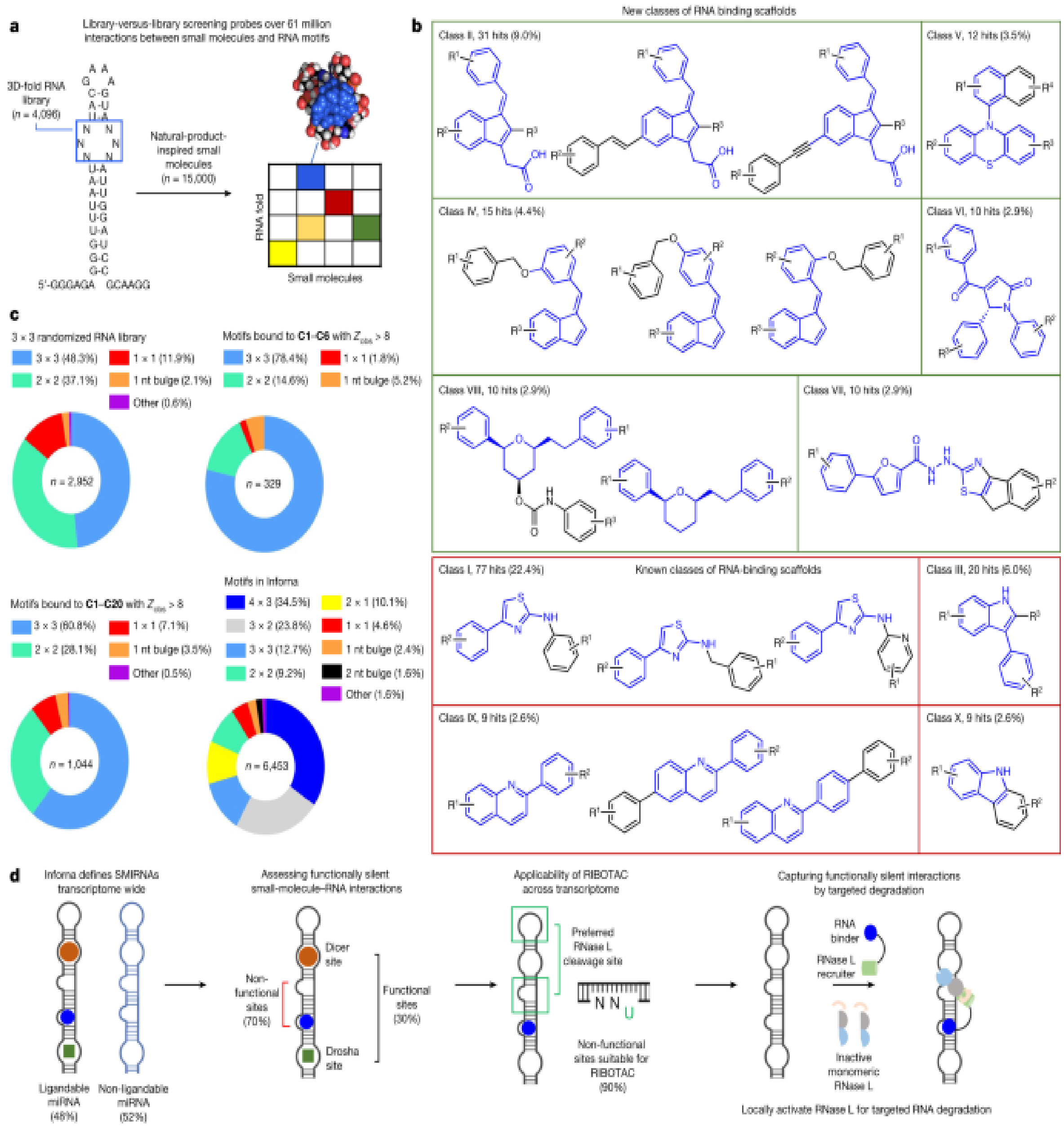
The importance of RNA in health and disease biology is well documented[2], affording opportunities within chemical biology to study function or intervene against dysfunction, respectively. Sequence-based targeting of RNA is often accomplished with complementary oligonucleotides that bind to and then recruit a ribonuclease to cleave the target[3]. This modality is best suited for targeting unstructured regions in an RNA, as molecular recognition occurs through base pairing[4]. However, RNA can be highly structured, and its biological function is often dictated by its structure[5,6]. Notably, such structured regions are amenable to targeting by the binding of small molecules that interact with pockets presented by an RNA fold[7]. However, the occupancy of RNA structures by small molecules alone is often not sufficient to elicit a biological effect[8].

Here we developed a strategy that converts biologically inactive RNA-binding small molecules into potent and specific effectors of function, achieved by affixing an RNA molecular recognition element to a second compound that binds to and activates a ribonuclease to cleave the target. Our focus is threefold: (1) define binding interactions between small molecules and RNA folds; (2) convert, in a programmable manner, highly selective binding interactions that are biologically inert into inducers of targeted degradation, affording potent and selective functional inhibitors; and (3) establish a paradigm by which small molecules can eliminate RNAs.

Identification of RNA-binding chemotypes

A 15,000-member, natural-product-like small molecule compound collection with diverse properties[9] was investigated for binding to a library of RNA 3D folds presented in a 3 × 3 internal loop library (ILL; 61,440,000 potential binding interactions probed) (Fig. 1a). The 4,096 unique RNA folds in 3 × 3 ILL include 1 × 1, 2 × 2 and 3 × 3 internal loops as well as bulge loops and fully paired RNAs. Binding was assessed using a fluorescent-dye-displacement assay[10], in which compounds from the library that bind to the RNA displace the dye and decrease its emission. A primary triage of the 15,000 small molecules (10 μM) yielded 1,584 hit compounds (Extended Data Fig. 1a). Secondary validation of the top 480 compounds yielded 344 binding small molecules. These diverse compounds include six RNA-binding scaffolds, such as 1-benzylidene-1-indene and phenothiazine (Fig. 1b).

Fig. 1: Library-versus-library screening defines new RNA-binding small molecules and druggable targets.



a, 2DCS analysis of more than 61 million theoretical interactions, identifying new interactions between small molecules and RNA motifs. **b**, The newly identified small-molecule RNA binders ($n = 344$) included 156 different scaffolds that fall into 79 major classes based on scaffold similarities. Among the top 10 most abundant classes (collectively covering 59.6% of all hits), six are new classes (green). **c**, Motif distribution from a 3 × 3 randomized RNA library used for 2DCS screening. As expected, 3 × 3 and 2 × 2 internal loops comprise the majority (85.4% total) of the library. Motifs that bound to **C1**–**C6** showed a significant enrichment for 3 × 3

internal loops ($P < 0.001$) and one-nucleotide bulges ($P < 0.001$). A total of 1,044 motifs bound to **C1**–**C20** with $Z_{obs} > 8$. Preference for 3 × 3 internal loops and one-nucleotide bulges was collectively observed for these compounds. Of these 1,044 motifs, only 23 (2.2%) are present in highly expressed human transcripts ($n = 2,712$ total motifs), and 375 are new motifs with no previously known small-molecule binder. Inforna contains over 100,000 RNA–small molecule interactions and 6,453 unique RNA motifs of various types. **d**, Although around 6% of all miRNAs can be bound by **C1**–**C6**, only about 30% of targetable sites within them are functional (Drosha or Dicer processing site) and are therefore predicted to induce a biological effect. The other approximately 70% are unproductive interactions that are predicted to be biologically silent. We identified that 48% of miRNAs that have ligandable non-functional sites are potential substrates for RNase L, which could be targeted by RIBOTACs. Thus, biologically inert binders can be converted into bioactive RIBOTACs that provoke targeted degradation. Statistical significance referred to in **c** was calculated using two-tailed Student's *t*-tests.

The binding of these 344 small molecules was studied using an orthogonal, microarray-based method named AbsorbArray[11], yielding 26 compounds (Extended Data Fig. 1b). Notably, AbsorbArray enables parallel affinity purification of the precise RNA members of the 3 × 3 ILL preferred by each small molecule when completed in the presence of excess unlabelled competitor oligonucleotides that mimic regions common to all library members, a selection method named two-dimensional combinatorial screening (2DCS)[12]. Of the 26 hit compounds from the primary and validation screens, six compounds (**C1**–**C6**) bound to the RNA library under these highly stringent conditions (Extended Data Fig. 1b and Supplementary Table 1). To determine the RNA 3D folds preferred by **C1**–**C6**, the frequency of each motif in the selected library was measured using RNA sequencing (RNA-seq) and compared to its frequency in the starting RNA pool. The statistical significance of the resulting enrichment was quantified by a pooled population comparison as Z_{obs}[13]. Collectively, these studies defined the affinity landscapes for the six small molecules, that is, structure–activity relationships (SARs) between small molecules and their preferred RNA 3D folds (Supplementary Table 1).

Three lines of evidence support that the six molecules identified here are new RNA-binding molecules. First, chemical similarity was assessed by calculating Tanimoto coefficients[14], where a threshold of <0.7 typically indicates dissimilar molecules[15]. The mean Tanimoto coefficient of the six small molecules identified here to all known RNA binders that are housed in two databases (Inforna, 404 small molecules[16]; and R-BIND, 104 small molecules[17]) was 0.25 ± 0.07 (Extended Data Fig. 1c,d). Second, a comparison of the physicochemical properties of **C1**–**C6** to known RNA-binding small molecules showed key differences in the total polar surface area (TPSA) as well as in the number of hydrogen bond donors and acceptors (Extended Data Fig. 1e). Finally, scaffold analysis revealed six new RNA-binding scaffolds, including azolium salts, bipyrrole pyrrolium salts and chromones (Extended Data Fig. 1d,f).

To define SAR, the RNA 3D folds significantly enriched for each small molecule with a $Z_{obs} > 8$—a statistical threshold considered to be bound by the small molecule[13]—were analysed ($n = 329$ unique motifs; Fig. 1c). Notably, these compounds collectively showed a preference for 3 × 3 internal loops (78.4% bound by small molecules versus 48.3% in the initial library; $P < 0.001$) and one-nucleotide bulges (5.2% compared to 2.1%; $P < 0.001$; Fig. 1c). Among the 329 unique motifs, 258 had no known small-molecule binder. LOGO analysis of the RNAs in the top 0.5% of statistically significant enriched loops ($Z_{obs} > 8$) revealed that each molecule binds to a unique RNA sequence pattern (Extended Data Fig. 2).

To generate a transcriptome-wide map of the RNA folds bound by **C1**–**C6**, we compared them to a database of folded human RNA structures[18]. The database includes structural information for all human primary (pri-) and precursor (pre-) microRNAs (miRNAs; 7,436 non-canonically paired motifs), and for highly expressed transcripts with a known structure, including 5S, 16S and 23S rRNAs, 7SL (signal-recognition particle), RNase P RNA, U4/U6 small nuclear RNA and 465 non-redundant tRNAs. Informatic mining of these structures against the RNA fold–small molecule interactions identified here showed that 13 motifs bound by **C1**–**C6** were present 114 times in 111 human miRNAs; that is, approximately 6% of human miRNAs have structures that are targetable by **C1**–**C6**. Of these targetable structures, only 33 (28.9%) are present in a functional Drosha or

Dicer processing site, indicating that most interactions would afford unproductive target engagement, that is, inactivity in a biological system (Fig. 1d). Notably, only 4 out of the 329 unique motifs identified from this selection (1.2%) bound by **C1**–**C6** were present in highly expressed human transcripts ($n = 2{,}712$ motifs).

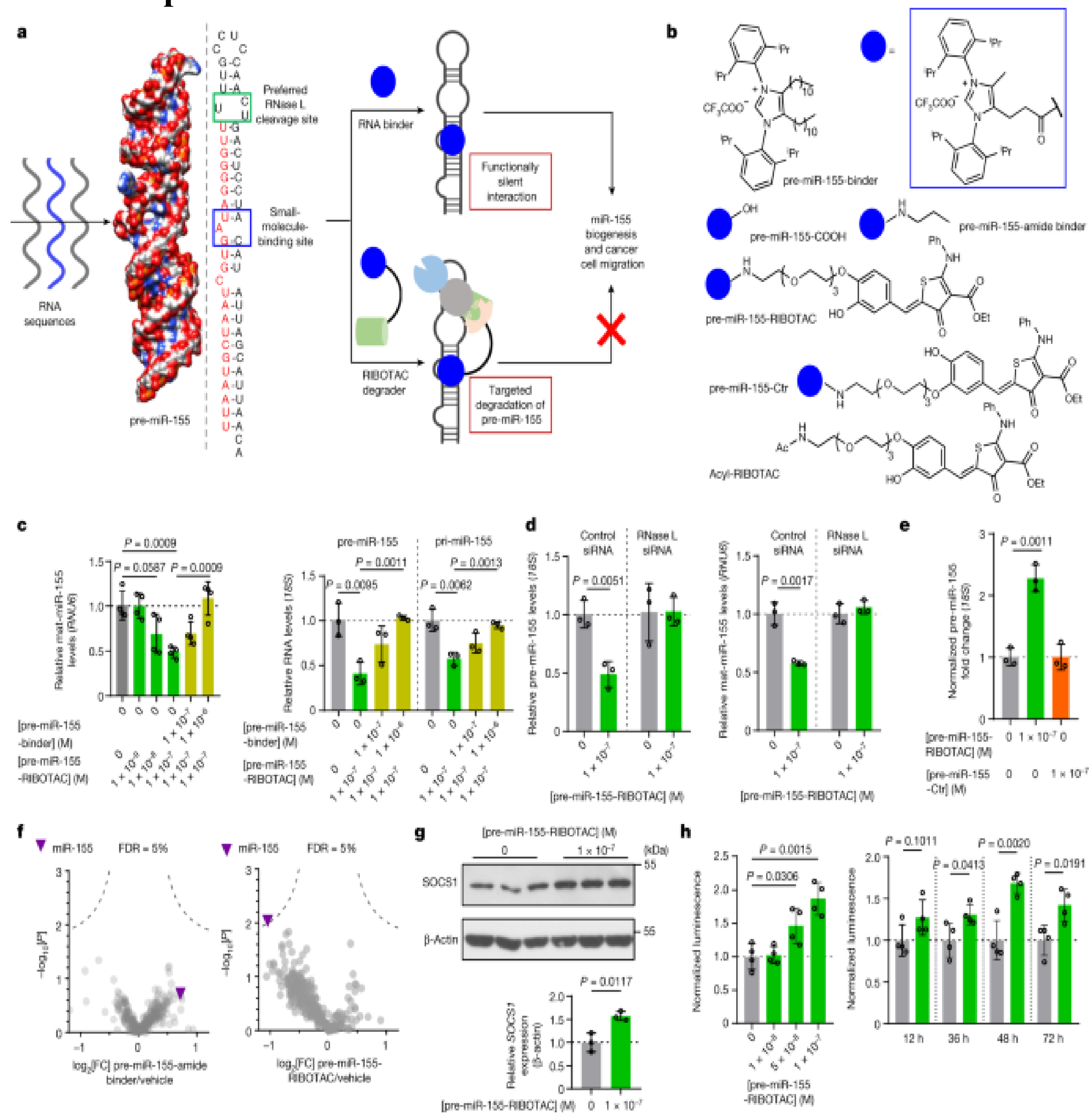
Design of a pre-miR-155 RIBOTAC degrader

Given the preponderance of biologically silent interactions, the challenge is therefore how to convert these into bioactive small molecules. One strategy is to imbue the binders with the ability to cleave its specific target, by generating a ribonuclease-targeting chimera (RIBOTAC)[1]. RIBOTACs are composed of an RNA-binding compound conjugated to a second small molecule that recruits and locally activates RNase L[1], which preferentially cleaves RNAs with a UNN pattern (with unpaired Us)[19,20,21,22] (Fig. 1d). Notably, of the predicted functionally silent interactions, 55 targetable miRNAs (48.2%) have at least one nearby RNase L site (within 10 bp) that could therefore be sensitive to a RIBOTAC (Fig. 1d).

One of the predicted biologically silent interactions identified is between **C1** (described previously[23,24]) and the 5′G**A**U/3′C_A bulge of pre-miR-155, where bold indicates an unpaired or non-canonically paired nucleotide (Fig. 2a). miR-155 has been implicated in inflammation[25], breast cancer[26] and other diseases. The **C1**–5′G**A**U/3′C_A interaction was identified from a member of the 3 × 3 ILL that is predicted to form two bulges, rather than an internal loop (Extended Data Fig. 3a). Furthermore, the 5′G**A**U/3′C_A binding site is proximal to a 5′U**U**U/3′G**UC**A motif that is predicted to be a substrate for RNase L[20,21] (Fig. 2a). Binding analysis by microscale thermophoresis (MST) on the 2DCS-selected RNA shows that **C1** binds only to the 5′GAU/3′C_A motif ($K_d = 1.1 \pm 0.2$ μM), similar to an RNA with both bulges ($K_d = 1.2 \pm 0.1$ μM) and a minimized RNA construct ($K_d = 0.8 \pm 0.1$ μM) (Extended Data Fig. 3b). By contrast, saturable binding was not observed to the RNA that displays only the other bulge or where a point mutation in the **C1**-targeting converted the A bulge into an AU pair (Extended Data Fig. 3a,b). These data were verified in an orthogonal binding

assay in which the A bulge was replaced with the fluorescent adenine mimic 2-aminopurine (2AP) ($K_{d,app} = 3.9 \pm 1$ μM) (Extended Data Fig. 3c).

Fig. 2: Pre-miR-155-RIBOTAC selectively degrades pre-miR-155 in an RNase-L-dependent manner in breast cancer cells.



a, Schematic of converting an inert binder engaging pre-miR-155 into an active RIBOTAC degrader. **b**, Structures of the compounds used to target pre-miR-155. **c**, The effects of pre-miR-155-RIBOTAC on mature (mat) (n = 4 biological replicates), pre- (n = 3 biological replicates) and pri- (n = 3 biological replicates) miR-155 levels, competed by increasing

concentrations of pre-miR-155-binder in MDA-MB-231 cells. **d**, The effect of siRNA knockdown of RNase L on pre-miR-155-RIBOTAC-mediated cleavage of pre-miR-155 in MDA-MB-231 cells, as determined using RT–qPCR. $n = 3$ biological replicates. **e**, Immunoprecipitation of pre-miR-155 using an anti-RNase L antibody in the presence of pre-miR-155-RIBOTAC in MDA-MB-231 cells ($n = 3$ biological replicates). **f**, The effect of pre-miR-155-amide-binder (left; 100 nM; $n = 4$ biological replicates) and pre-miR-155-RIBOTAC (right; 100 nM; $n = 3$ biological replicates) on the levels of the 373 miRNAs expressed in MDA-MB-231 cells. FC, fold change. **g**, Western blot analysis of SOCS1, a direct target of miR-155, after treatment of MDA-MB-231 cells with pre-miR-155-RIBOTAC ($n = 3$ biological replicates). **h**, The effect of pre-miR-155-RIBOTAC on the activity of a *SOCS1* 3′ UTR-luciferase reporter transfected into HEK293T cells, establishing both dose (left) and time dependence (right; $n = 4$ biological replicates). Data are mean ± s.d. (**c**–**e**, **g** and **h**). Statistical significance was determined using two-tailed Student's *t*-tests (**c**–**e**, **g** and **h**) or a Wald's test (**f**).

Source data

To our knowledge, azoliums have not been known to bind RNA. Thus, the RNA-binding ability of over 200 azolium derivatives was investigated. Of these compounds, only 14 (**C7**–**C20**) bound to RNA in the presence of competitor oligonucleotides (Supplementary Table 1). Notably, 10 of these 14 azolium salts share a cholesterol-derived sterol backbone, six of which are fused to the azolium ring. In total, 715 motifs are bound by **C7**–**C20** with $Z_{obs} > 8$, among which 117 are new with no previously known small-molecule binders (Extended Data Figs. 4 and 5 and Supplementary Table 1). Combining these with motifs bound by **C1**–**C6**, this study contributes 375 new RNA motifs to the current database of RNA–small-molecule interactions. An analysis of the RNAs selected by the azolium salts revealed that the top three small molecules predicted to bind to the 5′GAU/3′C_A motif in pre-miR-155 were **C19**, **C1** and **C20**, in rank order. Affinity measurements showed that only **C1** and **C20** bind to the A bulge of miR-155, while **C19** binding was undetermined due to aggregation under assay conditions (Extended Data Fig. 3d). As observed for **C1**, **C20** showed no saturable binding to an RNA with the A bulge changed to an AU pair (Extended Data Fig. 3d).

We first tested the hypothesis that **C1** (hereafter pre-miR-155-binder), engages the miR-155 precursor in cells but is biologically inert; that is, its binding to the miR-155 precursor did not inhibit biogenesis nor reduce mature miR-155 levels. To study target engagement, we used chemical cross-linking and isolation by pull-down (Chem-CLIP)[27]. In Chem-CLIP, an RNA-binding molecule is modified with a cross-linking moiety (chlorambucil (CA)) and a biotin purification module to enable pull-down with streptavidin beads (Extended Data Fig. 6a,b). To install the required modules, we hypothesized that the *n*-undecyl chains do not significantly contribute to binding. To test this, the *n*-undecyl chains were replaced with a propionic acid linker to afford compound pre-miR-155-COOH (Fig. 2b). We then amidated pre-miR-155-COOH with propylamine to mimic the reaction to generate the Chem-CLIP probe, yielding pre-miR-155-binder-amide (Fig. 2b), which binds with similar affinity to the A bulge as the pre-miR-155-binder does in both the MST- and 2AP-binding assays (Extended Data Fig. 3b,c), supporting our molecular-recognition hypothesis.

To confirm on-target binding in cells, a Chem-CLIP probe, pre-miR-155-Chem-CLIP (Extended Data Fig. 6a,b), was synthesized by installing chlorambucil and biotin through the carboxylic acid handle. A control Chem-CLIP probe was also synthesized that lacks the RNA-binding module (Ac-CA-Biotin, where Ac indicates acylated; Extended Data Fig. 6a). Pre-miR-155-Chem-CLIP (100 nM) enriched pre-miR-155 by around sevenfold in MDA-MB-231 cells, a triple-negative breast cancer cell line that aberrantly expresses miR-155[28], whereas no enrichment was observed for Ac-CA-Biotin (Extended Data Fig. 6c). Importantly, this pull-down and enrichment of pre-miR-155 by pre-miR-155-Chem-CLIP was competed dose-dependently by co-treatment with **C1** (Extended Data Fig. 6c). Collectively, these studies demonstrate that **C1** directly binds to pre-miR-155 in cells. To affirm the precise binding site of **C1** within the miR-155 precursor, we used chemical cross-linking and isolation by pull-down to map small molecule RNA-binding sites (Chem-CLIP–Map-Seq)[29]. Indeed, in vitro, pre-miR-155-Chem-CLIP cross-linked adjacent to the A bulge-binding site (Extended Data Fig. 6d).

The effects of the pre-miR-155-binder on miR-155 biogenesis was assessed by measuring levels of pre- and mature miR-155 using quantitative PCR

with reverse transcription (RT–qPCR). As expected, no effect was observed on the levels of either form of the miRNA from 0.01 to 5 μM, concentrations at which no toxicity was observed; similarly, pre-miR-155-amide-binder was also inactive (Extended Data Fig. 6e). Notably, the 5′GAU/3′C_A bulge found in the miR-155 precursor is also found in 11 other human miRNAs, including three miRNAs in which the A bulge is present in the Dicer processing site (miR-18a, miR-196a and miR-4435). Indeed, Chem-CLIP studies revealed that 4 out of these 11 miRNAs were directly engaged in MDA-MB-231 cells—miR-1226, miR-3168, miR-4435 and miR-4700 (Extended Data Fig. 6f). Note that six of the remaining miRNAs were not detectable in this assay, and pre-miR-4640 was not enriched. As expected, because the small molecule binding site is also a processing site, pre-miR-155-amide-binder reduced the levels of mature miR-18a, miR-196a and miR-4435 in a dose-dependent manner, but did not affect the levels of the other nine miRNAs (Extended Data Fig. 6f).

Given thatpre-miR-155-binder engages the pre-miR-155 target yet is biologically inert, we converted this RNA binder into a RIBOTAC through conjugation of pre-miR-155-COOH to an RNase-L-recruiting small molecule[1], affording pre-miR-155-RIBOTAC (Fig. 2b). Approaches for converting functionally silent molecules into functional chimeric molecules are well studied in the protein-degradation field[30,31,32], but not in the RNA-degradation field. Two control compounds were also synthesized—Ac-RIBOTAC, the RNase L-recruiting module lacking pre-miR-155-binder, and pre-miR-155-Ctr, which contains a less active RNase-L-recruiting module (Fig. 2b). Notably, the affinity of pre-miR-155-RIBOTAC for 5′GAU/3′C_A using the minimized binding site is similar to that of pre-miR-155 binder, whether measured by MST (K_d = 0.8 ± 0.1 μM versus 0.53 ± 0.09 μM, respectively) or using the 2AP-substituted bulge (K_d = 3.9 ± 1.2 μM versus 4.6 ± 1.2 μM, respectively) (Extended Data Fig. 3b,c). Furthermore, pre-miR-155-RIBOTAC binds to pre-miR-155 with an affinity of 2.2 ± 0.9 μM in the 2AP assay with no binding observed for a pre-miR-155 mutant in which the A bulge is mutated to an AU pair (Extended Data Fig. 3e).

In vitro, pre-miR-155-RIBOTAC recruited RNase L to pre-miR-155 and induced its cleavage at the predicted cleavage site—the 5′U**U**U/3′G**UC**A motif—in a dose-dependent manner (Extended Data Fig. 7a). Furthermore,

cleavage was reduced, dose-dependently, after addition of increasing concentrations of pre-miR-155 binder (Extended Data Fig. 7a). Importantly, mutation of the **C1**-binding site or of the cleavage site abolished the activity of pre-miR-155-RIBOTAC, indicating that both sites are required for cleavage (Extended Data Fig. 7b). Similarly, Ac-RIBOTAC did not induce cleavage of pre-miR-155 (Extended Data Fig. 7b). These in vitro results, for both pre-miR-155-RIBOTAC and pre-miR-155-Ctr, were verified using a model of the pre-miR-155 A bulge labelled at the 5′ and 3′ ends with fluorescein and a black hole quencher, respectively (Extended Data Fig. 3f).

As in vitro studies suggested specific degradation of pre-miR-155 through the recruitment of RNase L by pre-miR-155-RIBOTAC, its effect on pri-, pre- and mature miR-155 levels was studied in MDA-MB-231 cells. Treatment with 100 nM of pre-miR-155-RIBOTAC reduced mature miR-155 levels in a time-dependent manner with no observed toxicity and a maximum reduction of 71 ± 10% after 48 h (Extended Data Fig. 7c,d and Supplementary Figs. 2 and 3). Furthermore, a dose-dependent reduction in all three species was observed after a 48 h treatment period, and each was rescued by co-treatment with pre-miR-155-binder (Fig. 2c), indicating that the two molecules compete for the same binding site. Moreover, removal of pre-miR-155-RIBOTAC in a wash-out study resulted in a time-dependent restoration of pre-miR-155 levels with a half-life of about 19.3 h (Extended Data Fig. 7e), in accordance with the reported turnover of miR-155 in cells[33]. Neither control molecule Ac-RIBOTAC nor pre-miR-155-Ctr affected pre- or mature miR-155 levels in cells (Extended Data Fig. 7f). Similarly, in CFPAC-1 cells—a pancreatic cancer cell line that overexpresses miR-155[34]—pre-miR-155-RIBOTAC reduced precursor and mature levels of miR-155 by around 20% and 65% respectively, whereas the pre-miR-155-binder-amide did not do so (Extended Data Fig. 7g).

To validate the mode of action for pre-miR-155-RIBOTAC, MDA-MB-231 cells were transfected with a short interfering RNA (siRNA) targeting RNase L or a control siRNA. Notably, RNase-L-siRNA ablated the ability of pre-miR-155-RIBOTAC to reduce miR-155 levels (Fig. 2d); compound activity is therefore dependent on this RNase. Furthermore, RNase L immunoprecipitation enriched pre-miR-155 by around 2.4-fold from MDA-MB-231 cells treated with pre-miR-155-RIBOTAC, but not pre-miR-155-Ctr, both as compared with treatment with vehicle (Fig. 2e). Thus, pre-miR-

155-RIBOTAC directly engages the miR-155 precursor and recruits RNase L to cleave the transcript.

pre-miR-155-RIBOTAC is selective

The selectivity of pre-miR-155-RIBOTAC for miR-155 was assessed miRNome-wide using RT–qPCR profiling. An analysis of 373 miRNAs expressed in MDA-MB-231 cells showed that pre-miR-155-RIBOTAC selectively reduced miR-155 levels in a similar manner as LNA-155 (Extended Data Fig. 7h), without significant effects on any other miRNAs (Fig. 2f), including those that have the same A bulge. Note that repression of a subset of these miRNAs was observed at pre-miR-155-amide-binder concentrations of higher than 1 µM, tenfold greater than the concentration of pre-miR-155-RIBOTAC (100 nM) (Extended Data Fig. 6f). To confirm these findings, the selectivity of pre-miR-155-RIBOTAC and LNA-155, an antagomir that targets mature miR-155-5p, was also assessed using miRNA-seq. Notably, both modalities have a similar effect across the miRNome (Extended Data Fig. 7h).

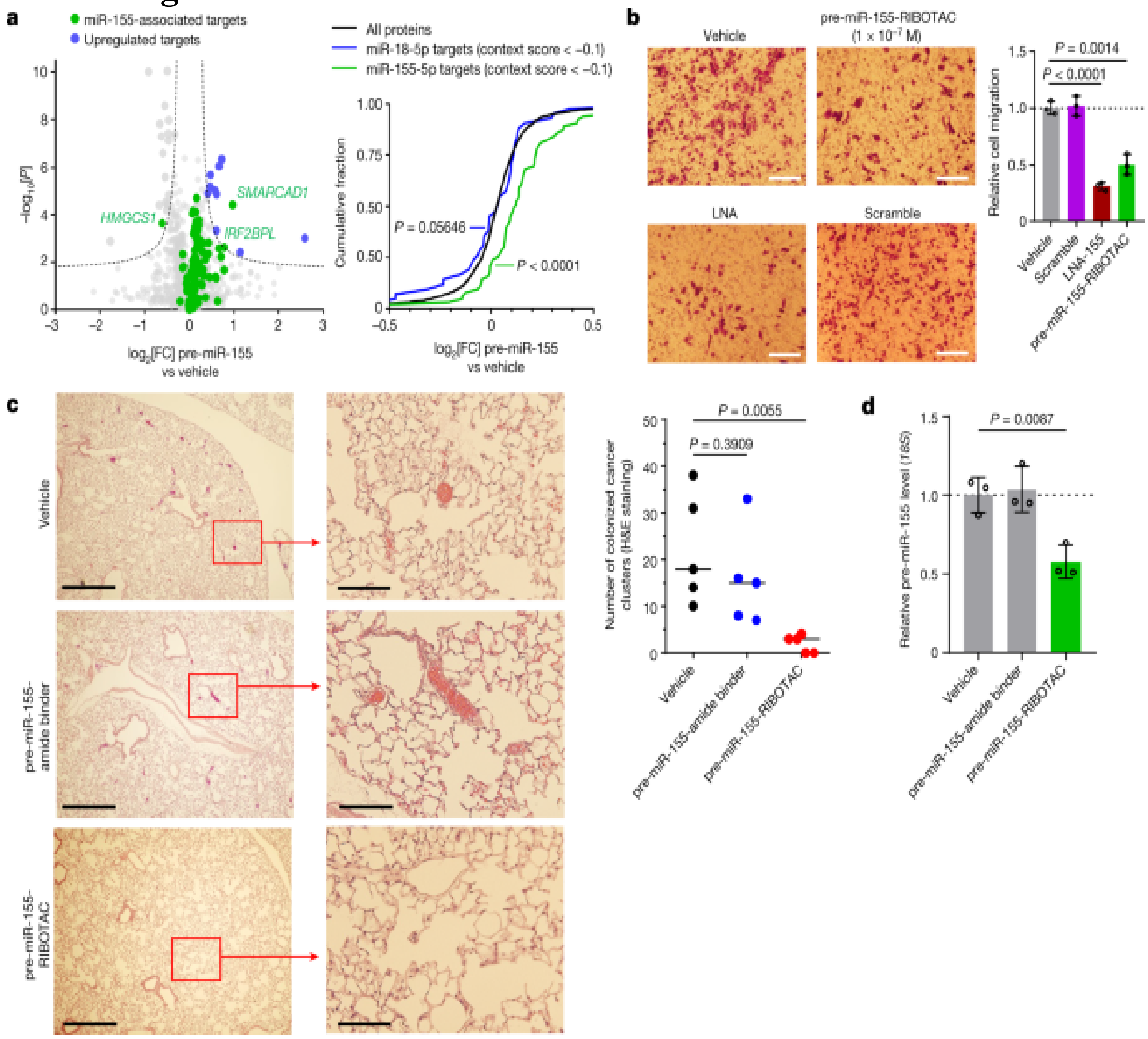
The miRNAs with the same A bulge—miR-18a, miR-101-1, miR-1226 (bound by **C1** in Chem-CLIP studies), miR-3945 and miR-4435 (bound by **C1** in Chem-CLIP studies)—also have potential RNase cleavage sites (Extended Data Fig. 6f). However, it appears that the miR-155 precursor harbours the superior RNase L cleavage site. In vitro, cleavage occurred at nucleotides U28–U30, within and adjacent to a 1 × 2 asymmetric loop juxtaposed to the apical hairpin loop. RNA secondary structure prediction indicates that this loop is probably dynamic and may form a large hairpin of single-stranded nucleotides with only a 1 kcal mol^{-1} difference in predicted free energy. Furthermore, the RNase L cleavage sites in these five other miRNAs are at different distances from the binding site than in miR-155 (7 bp), suggesting that distances between binding and degrader sites can be used to program selectivity (Extended Data Fig. 6f). Collectively, these studies show that both the RNA-binding site and the RNase L cleavage site are required for pre-miR-155-RIBOTAC to induce degradation of its target RNA.

To assess the selectivity of pre-miR-155-RIBOTAC on the transcriptome, RNA-seq analysis was performed in MDA-MB-231 cells treated with either pre-miR-155-RIBOTAC or LNA-155 (Extended Data Fig. 7i). Among the 13,332 transcripts detected, 29 (0.22%) were significantly affected by pre-miR-155-RIBOTAC treatment ($P < 0.05$ and $\log_2$[fold change] > 1), with 1 upregulated and 28 downregulated. Notably, these 29 transcripts were also affected to a comparable extent by LNA-155 treatment (Extended Data Fig. 7i). When considering downstream targets of miR-155 predicted by TargetScanHuman (v.7.0)[35] ($n = 469$), 307 (65%) were upregulated by pre-miR-155-RIBOTAC. A similar percentage (68%) of these 469 targets was also upregulated by LNA-155. Notably, 263 targets were upregulated by both pre-miR-155-RIBOTAC and LNA-155. A comparison of the normalized read counts for all genes between pre-miR-155-RIBOTAC and LNA-155 treatment showed a highly significant correlation, with $R > 0.99$ (Extended Data Fig. 7i). Collectively, these data indicate that pre-miR-155-RIBOTAC affects the transcriptome in similar ways to an oligonucleotide targeting miR-155 (LNA-155) and with limited off-target effects, although additional studies are needed to assess the selectivity of miRNA knockdown. Notably, pre-miR-155-RiboTAC selectively reduced miR-155 levels miRnome-wide in MDA-MB-231 cells forced to express wild-type pre-miR-155 but not those forced to express a binding site mutant (Extended Data Fig. 8a).

Upregulation of miR-155 modulates breast cancer cell migration and invasion by repressing SOCS1, which dampens cytokine signalling[26]. Indeed, pre-miR-155-RIBOTAC (100 nM) treatment increased SOCS1 levels by around 50% (Fig. 2g). These data were confirmed using a luciferase reporter fused to the SOCS1 3′ untranslated region (UTR), establishing both dose and time dependence of the RIBOTAC's derepression of miR-155's direct target (Fig. 2h). Global proteomics analysis confirmed that selective effects were induced by pre-miR-155-RIBOTAC. Of the 3,158 detectable proteins, 385 were upregulated and 98 of these are associated with miR-155 (Fig. 3a). Among these 98 proteins, the transcripts levels of 86 (88%) were also upregulated as determined using RNA-seq analysis. Notably, although SOCS1 was not detectable, the change in two miR-155-regulated proteins, IRF2BPL and SMARCAD1, was statistically significant ($P < 0.01$). Finally, inspection of proteins associated with pre-miR-18a,

which was not affected by pre-miR-155-RIBOTAC, revealed that those that were detectable ($n = 42$ out of 264) were unaffected by treatment with this RIBOTAC (Fig. 3a).

Fig. 3: Pre-miR-155-RIBOTAC selectively degrades pre-miR-155 and reduces lung colonization in vivo.



a, Left, proteome-wide changes in MDA-MB-231 cells treated with pre-miR-155-RIBOTAC (100 nM) versus vehicle ($n = 3$ biological replicates). Right, pre-miR-155-RIBOTAC significantly upregulated miR-155 related proteins ($n = 98$ proteins), as indicated by a Kolmogorov–Smirnov analysis (right) of their levels versus all proteins ($n = 3$ biological replicates). **b**, The effect of pre-miR-155-RIBOTAC on MDA-MB-231 cell migration ($n = 3$ biological replicates); 2 fields of view were quantified per replicate.

Scale bars, 0.5 mm. **c**, pre-miR-155-RIBOTAC suppresses lung colonization in vivo, as determined by counting lung nodules ($n = 5$ mice) and by haematoxylin and eosin (H&E) staining ($n = 5$ mice; 2 fields of view were quantified per replicate). Scale bars, 1 mm (left) and 0.2 mm (right). **d**, The effect of pre-miR-155-amide-binder and pre-miR-155-RIBOTAC on pre-miR-155 levels in vivo, as determined by RT–qPCR using primers selective for human pre-miR-155 ($n = 3$ mice). Data are mean ± s.d. (**b** and **d**). Statistical significance was determined using a Wald's test (**a**) or two-tailed Student's *t*-tests (**b**–**d**).

Source data

The cleavage of pre-miR-155 and subsequent upregulation of SOCS1 suggested that pre-miR-155-RIBOTAC may inhibit the migration of MDA-MB-231 cells. Indeed, treatment with pre-miR-155-RIBOTAC, but not with pre-miR-155-Ctr, reduced MDA-MB-231 cell migration by around 50% (Fig. 3b and Extended Data Fig. 8b). To confirm that these effects of pre-miR-155-RIBOTAC are due to abrogation of the miR-155 circuit, MCF-10a cells—a model of normal breast epithelial cells—were engineered to express either wild-type pre-miR-155 or a mutant in which the **C1**-binding site was abolished by mutation of the A bulge to a base pair (Extended Data Fig. 9a). Forced expression of both pre-miRNAs increased the migratory ability of MCF-10a cells (Extended Data Fig. 9b), indicating that miR-155 indeed contributes to this phenotype as previously reported[28,36]. Notably, treatment of MCF-10a cells expressing WT, but not mutant, pre-miR-155 with pre-miR-155-RIBOTAC reduced pre-miR-155 levels as well as the migration of these cells (Extended Data Fig. 9a–c).

Complementary analysis of pre-miR-155-RIBOTAC was performed in human umbilical vein endothelial cells (HUVECs), in which miR-155 controls angiogenesis through modulation of the Von Hippel Lindau (VHL) protein[37]. Treatment with pre-miR-155-RIBOTAC (100 nM) downregulated miR-155 levels to a similar extent to that observed in MDA-MB-231 cells, whereas pre-miR-155-binder and pre-miR-155-Ctr were inactive (Extended Data Fig. 9d–f). Moreover, there were no effects of pre-miR-155-RIBOTAC on the levels of other miRNAs with pre-miR-155-binder-binding sites (Extended Data Fig. 9g). Finally, reduction of miR-155 by pre-miR-155-RIBOTAC boosted VHL protein levels by around 50% and reduced tubule

branching by approximately 35%, indicating reduced angiogenic ability (Extended Data Fig. 9h,i).

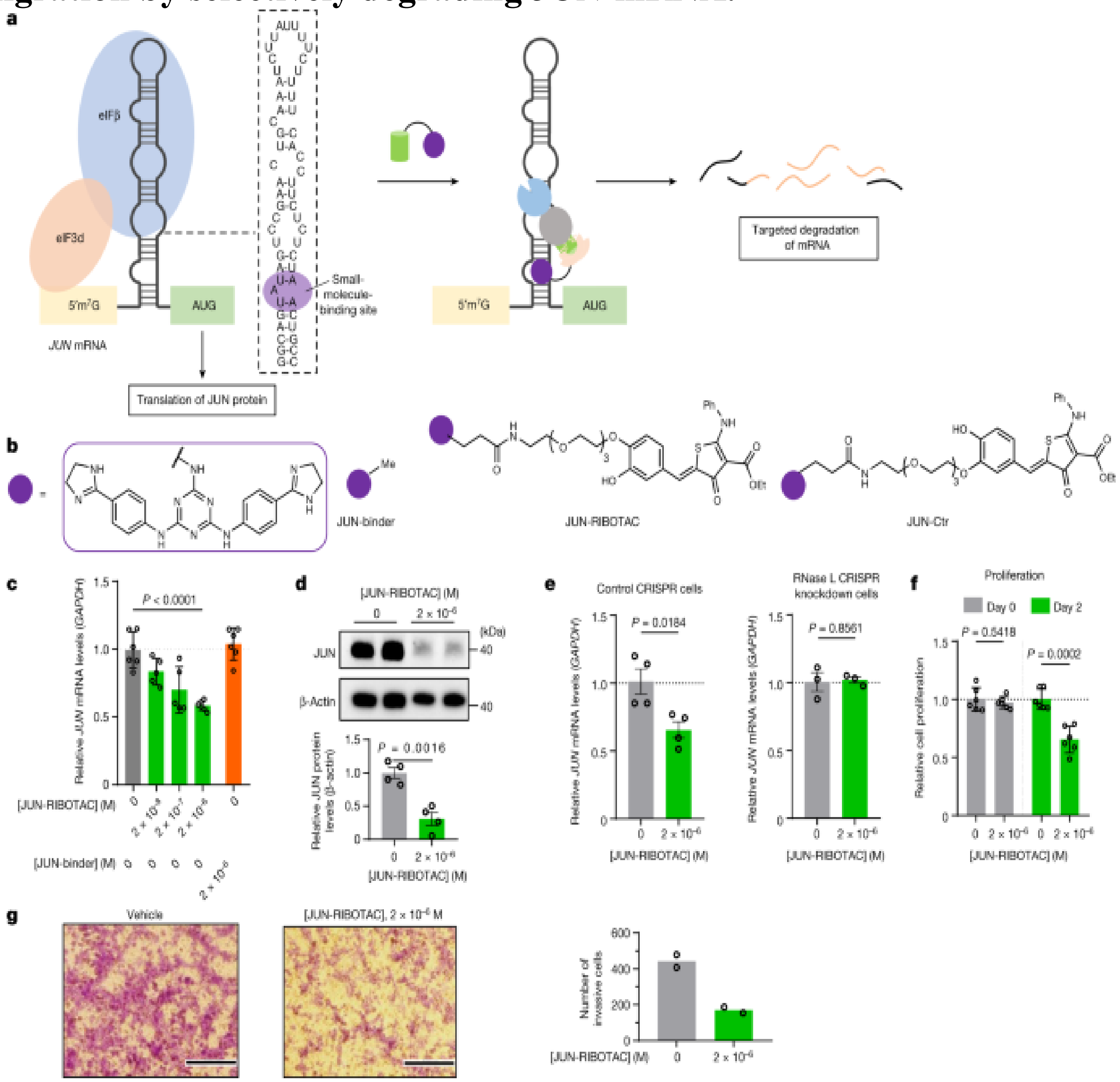
To assess the ability of pre-miR-155-RIBOTAC to inhibit miR-155-mediated colonization of breast cancer to the lungs, mice were injected through the tail vein with MDA-MB-luc cells[38]. After 5 days, mice were treated with pre-miR-155-RIBOTAC (1 mg per kg, every other day) by intraperitoneal injection. The dosage was determined by drug metabolism and pharmacokinetics studies (C_{max} = 1.2 µM in the plasma with half-life of 1.8 h; Extended Data Fig. 8c). As evidenced by the reduced number of nodules, lung colonization was significantly inhibited in the pre-miR-155-RIBOTAC-treated group versus the vehicle- or pre-miR-155-binder-amide-treated cohorts (Fig. 3 and Extended Data Fig. 8d). Histological staining of lung tissue sections showed marked reductions in tumour burden after treatment with the RIBOTAC (Fig. 3c). As mouse pre-miR-155 has a different secondary structure lacking the binding site for pre-miR-155-RIBOTAC, fluorescence in situ hybridization (FISH) and RT–qPCR were used to assess the levels of human mature and pre-miR-155, respectively. There were marked reductions in mature (Extended Data Fig. 8e) and pre-(Fig. 3d) miR-155 in human TBNC cells in mouse lung tissues after treatment with pre-miR-155-RIBOTAC, whereas treatment with pre-miR-155-binder-amide had no significant effect. Collectively, these results demonstrated that pre-miR-155-RIBOTAC can impair miR-155-mediated tumour colonization in vivo.

RIBOTACs targeting *JUN* and *MYC* mRNAs

To test whether our approach to convert biologically silent interactions into bioactive ones is broadly applicable, we designed RIBOTACs for two other oncogenic RNAs—*JUN* and *MYC*. Notably, both oncoproteins are intrinsically disordered and are considered to be undruggable. Their transcripts harbour an internal ribosomal entry site (IRES) in their 5′ untranslated regions (UTRs) that can drive cap-independent translation of MYC[39,40] and cap-dependent specialized translation initiation of JUN[41,42]. To determine whether these mRNAs form targetable structures in their 5′ UTRs, we coupled Inforna with the state-of-the-art RNA structure computing program, ScanFold[43], which identifies regions of unusual

thermodynamic stability within an RNA, an indicator of function and potential evolutionary conservation. Indeed, both mRNAs form a thermodynamically stable region in their IRESs that house structures that are targetable with small molecules and that are proximal to potential RNase-L-cleavage sites (Figs. 4a and 5a).

Fig. 4: JUN-RIBOTAC impairs pancreatic tumour cell proliferation and migration by selectively degrading *JUN* mRNA.

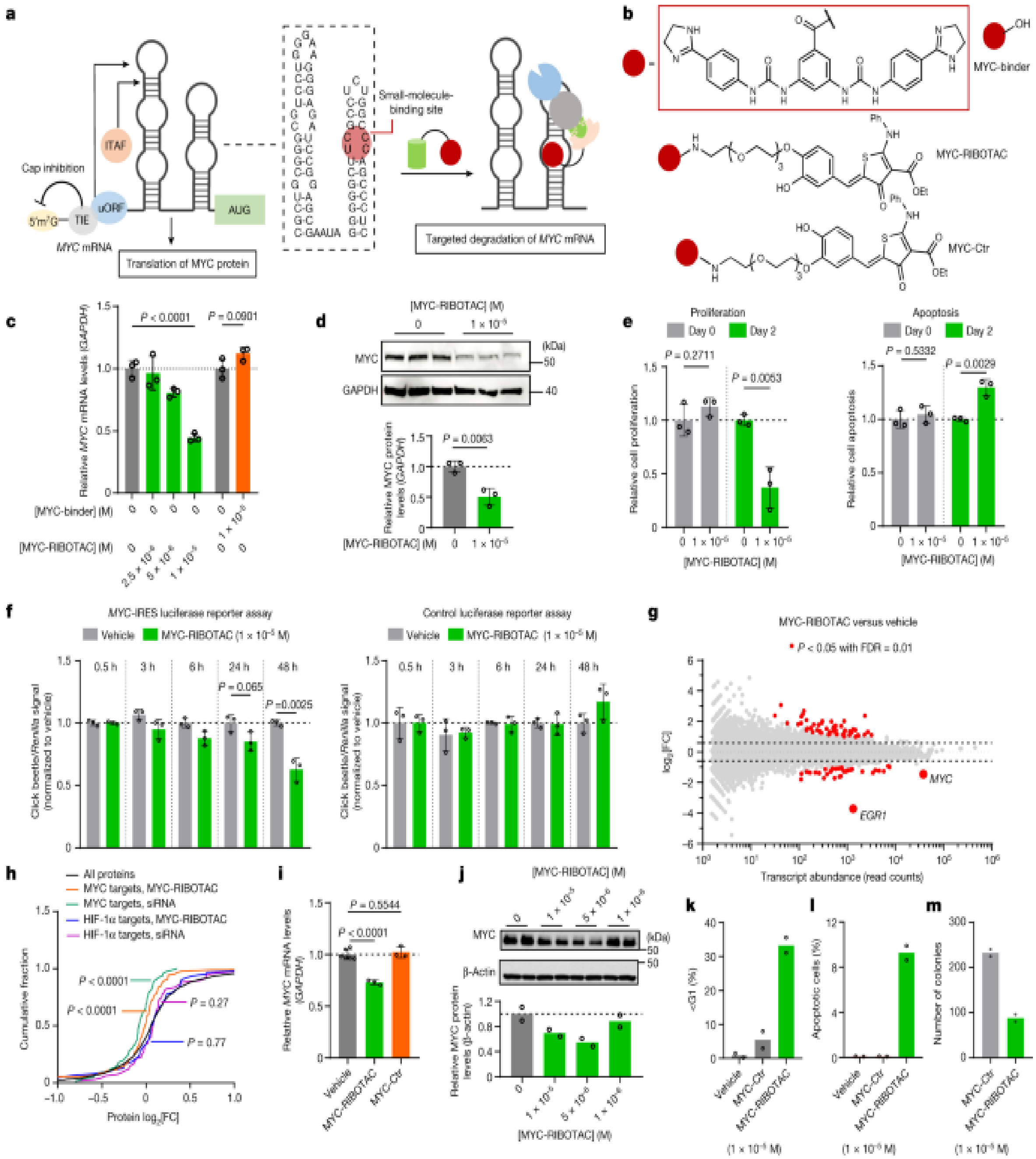


a, Schematic of *JUN* degradation by targeting the *JUN* IRES. **b**, The structures of compounds used to target *JUN* mRNA. **c**, The effect of JUN-RIBOTAC and JUN-binder on *JUN* mRNA levels in MIA PaCa-2 cells after treatment for 72 h, as determined using RT–qPCR ($n = 6$ biological

replicates). **d**, The effect of JUN-RIBOTAC on JUN protein levels in MIA PaCa-2 cells ($n = 4$ biological replicates). **e**, The effect of JUN-RIBOTAC on *JUN* mRNA levels in MIA PaCa-2 cells in which RNase L was knocked down by CRISPR ($n = 3$ biological replicates) and in the corresponding MIA PaCa-2 control cell line in which CRISPR editing was performed using a scrambled guide RNA ($n = 4$ biological replicates), as determined using RT–qPCR. **f**, The effect of JUN-RIBOTAC on the proliferation of MIA PaCa-2 cells ($n = 6$ biological replicates). **g**, The effect of JUN-RIBOTAC on the invasiveness of MIA PaCa-2 cells, as determined using a Boyden chamber assay ($n = 2$ biological replicates; 2 fields of view were quantified per replicate). Data are mean ± s.d. (**c**–**f**). Statistical significance was determined using two-tailed Student's *t*-tests (**d**–**f**) and one-way analysis of variance (ANOVA) adjusted for multiple comparisons (**c**).

Source data

Fig. 5: MYC-RIBOTAC selectively targets MYC in an RNase-L-dependent manner.



a, Schematic of the targeted degradation of the *MYC* IRES. **b**, Compound structures. **c**, The effect of MYC-binder and MYC-RIBOTAC on *MYC* mRNA levels in HeLa cells, as determined using RT–qPCR. $n = 3$ biological replicates. **d**, The effect of MYC-RIBOTAC on MYC protein levels in HeLa cells ($n = 3$ biological replicates). **e**, The effect of MYC-RIBOTAC on the proliferation (left) and apoptosis (right) of HeLa cells ($n = 3$ biological replicates). **f**, The effect of MYC-RIBOTAC on *MYC* IRES luciferase reporter in HEK293T cells (left) or on a control reporter lacking the IRES

(right)($n = 3$ biological replicates). **g**, Transcriptome-wide changes in HeLa cells treated with MYC-RIBOTAC (10 μM) after treatment for 48 h ($n = 3$ biological replicates). *EGR1* is a well-known downstream target of MYC[50]. **h**, Cumulative distribution analysis of the effect of MYC-RIBOTAC and a *MYC*-selective siRNA on 87 well-validated downstream targets of MYC, or on the downstream targets of HIF-1α, as indicated by a Kolmogorov–Smirnov analysis of their levels relative to all proteins ($n = 3$ biological replicates). **i**, The effect of MYC-Ctr and MYC-RIBOTAC on *MYC* mRNA levels in Namalwa Burkitt lymphoma cells ($n = 3$ biological replicates) compared with the vehicle ($n = 6$ biological replicates). **j**, The effect of MYC-RIBOTAC on MYC protein levels in Namalwa cells ($n = 2$ biological replicates). **k**, The effect of MYC-RIBOTAC on the cell cycle of Namalwa cells. $n = 2$ biological replicates. **l**, The ability of MYC-RIBOTAC or MYC-Ctr to induce apoptosis of Namalwa cells ($n = 2$ biological replicates). **m**, The effect of MYC-RIBOTAC on colony formation of Namalwa cells ($n = 2$ biological replicates). Data are mean ± s.d. (**c**–**f** and **i**). Statistical significance was determined using a one-way ANOVA adjusted for multiple comparisons (**c**), two-tailed Student's *t*-tests (**d**–**i**), Wald's test (**g**), or Kolmogorov–Smirnov test (**h**).

Source data

The *JUN* IRES comprises a hairpin with various internal loops and bulges embedded in its stem that is recognized by eukaryotic initiation factor 3 (eIF3) (Fig. 4a). Inforna identified a small-molecule candidate for the A bulge (5′U**A**U/3′A_A), named JUN-binder (Fig. 4b). Notably, this A bulge is adjacent to a pyrimidine-rich internal loop (5′G**UCC**G/3′C**UCU**C), a probable substrate for RNase L. Using the 2AP-binding assay, JUN-binder binds to this loop with a half-maximal effective concentration (EC_{50}) of 1.1 ± 0.2 μM, with no observable binding to a control RNA in which the bulged nucleotide was converted to a base pair ($EC_{50} > 50$ μM) (Extended Data Fig. 10a). These data were corroborated by an orthogonal binding assay using Cy-5-labelled RNAs (Extended Data Fig. 10b). Target engagement was validated both in vitro and in the pancreatic cancer cell line MIA PaCa-2 (Extended Data Fig. 10c,d) using Chem-CLIP and its competitive variant. Finally, although in vitro binding assays and target validation confirmed engagement of the *JUN* IRES with JUN-binder, this molecule was

biologically inert, with no effects on *JUN* mRNA or protein levels up to concentrations of 2 μM (Fig. 4c and Extended Data Fig. 10e).

Given these findings, the RIBOTAC strategy was applied, affording JUN-RIBOTAC and the corresponding control molecule with the less active RNase L-recruiting regioisomer, JUN-Ctr (Fig. 4b). The binding affinities of these molecules were studied in the same manner as for the JUN-binder, using 2AP- or Cy5-labelled RNAs. JUN-RIBOTAC and JUN-Ctr bind to the *JUN* IRES similarly in the Cy5 assay with K_d values of 0.5 ± 0.2 μM and 0.4 ± 0.1 μM, respectively, whereas there was no saturable binding to a fully paired control RNA (Extended Data Fig. 10b). To test whether JUN-RIBOTAC induced cleavage of the *JUN* IRES by recruiting RNase L, the *JUN* IRES was labelled with a fluorophore and quencher at its 5′ and 3′ ends, respectively (Extended Data Fig. 10f). JUN-RIBOTAC increased fluorescence intensity, suggesting its ability to induce RNase L cleavage of the *JUN* IRES, while having no effect on a mutant in which the A bulge was converted to an AU pair; furthermore, JUN-Ctr did not induce cleavage of the wild-type IRES (Extended Data Fig. 10f).

In MIA PaCa-2 cells, JUN-RIBOTAC reduced *JUN* mRNA levels in a dose-dependent manner up to about 40% and reduced protein levels by around 75% at a dose of 2 μM (Fig. 4c,d and Supplementary Fig. 4), whereas JUN-binder was inactive (Extended Data Fig. 10e). To confirm the mode of action, we evaluated the activity of JUN-RIBOTAC in MIA PaCa-2 cells in which RNase L was knocked down by CRISPR (Supplementary Fig. 5). As predicted, JUN-RIBOTAC was inactive in the RNase L CRISPR cell line, whereas it showed around a 40% decrease in *JUN* mRNA level in the CRISPR control cell line (Fig. 4e). JUN contributes to proliferative and invasive phenotypes in different tumour types, including pancreatic cancer[44,45,46]. The reduction in JUN protein levels by JUN-RIBOTAC in MIA PaCa-2 cells was sufficient to inhibit both cell proliferation (~40% at 2 μM; Fig. 4f) and invasion (~60% at 2 μM; Fig. 4g).

The oncogenic transcription factor MYC coordinately regulates the cell cycle, proliferation and metabolism in many cancer types, including in cervical cancer as exemplified by HeLa cells[47,48]. Notably, the *MYC* IRES also forms a thermodynamically stable structure comprising two hairpins

and three internal loops (Fig. 5a)[19,20,21,22] and, fortuitously, one internal loop, 5′U**U**C**G**/3′A**C**C**C**, is targetable with a molecule from our database, named MYC-binder (Fig. 5b). Moreover, this loop is adjacent to a pyrimidine-rich hairpin loop (Fig. 5a). The phenylene bis-phenylurea group of MYC-binder interacts with the internal loop with a K_d of 2.3 ± 1 μM, as determined using a fluorescence-based binding assay with a Cy5-labelled RNA; no binding was observed to a fully paired RNA (Extended Data Fig. 11a). Furthermore, the derived MYC-Chem-CLIP probe engaged endogenous *MYC* mRNA in HeLa cells in a dose-dependent manner (about 50% enrichment at 20 μM; Extended Data Fig. 11b), and this enrichment was depleted by addition of competing concentrations of MYC-binder (Extended Data Fig. 11b). Finally, although the *MYC* IRES is a functional site, engagement by MYC-binder did not affect MYC protein levels or *MYC* mRNA levels (Fig. 5c and Extended Data Fig. 11c).

As MYC-binder was biologically inert, it was appended with the heterocyclic RNase-L-recruiter to afford MYC-RIBOTAC, or with the less active regioisomer to afford MYC-Ctr (Fig. 5b). MYC-RIBOTAC and MYC-Ctr bind to the *MYC* 5′U**U**C**G**/3′A**C**C**C** internal loop similarly in the Cy5 assay with K_d values of 1.1 ± 0.1 μM and 0.9 ± 0.2 μM, respectively, whereas no saturable binding was observed to a fully paired control RNA (Extended Data Fig. 11a). Akin to the dual-labelled *JUN* IRES model (Extended Data Fig. 9f), an in vitro fluorescent reporter was created to study whether MYC-RIBOTAC could recruit RNase L and induce *MYC* IRES degradation. Indeed, MYC-RIBOTAC but not MYC-Ctr recruited RNase L to the *MYC* IRES, as evidenced by the resulting increase in fluorescence, and again no change in fluorescence was observed after incubation of the fully paired RNA with either molecule (Extended Data Fig. 11d). Notably, pre-miR-155-RIBOTAC had no effect on the *MYC* IRES RNA and vice versa (Extended Data Fig. 11d), supporting the selectivity of the RIBOTAC approach with different RNA-binding modules.

In contrast to MYC-binder, MYC-RIBOTAC decreased the abundance of *MYC* mRNA in HeLa cells in a dose-dependent manner, up to around 50% at a 10 μM dose (Fig. 5c and Supplementary Figs. 6 and 7) with a concomitant reduction in MYC protein levels (Fig. 5d). A reduction in MYC protein was associated with reduced HeLa cell proliferation and the

induction of apoptosis (Fig. 5e). Notably, MYC-Ctr and two control molecules in which the RNase-L-recruiting module (Ac-RIBOTAC) or the RNA-binding module was conjugated to propylamine (MYC-amide-binder) were inactive (Extended Data Fig. 11e,f). To confirm that reduction of *MYC* mRNA levels by MYC-RIBOTAC was dependent on RNase L, we knocked down RNase L in HeLa cells with an siRNA. As expected, the effect of MYC-RIBOTAC on *MYC* mRNA levels was ablated in cells after RNase L knockdown (Extended Data Fig. 11g). Notably, the proteolysis-targeting chimera (PROTAC) MZ-1 targets the protein bromodomain containing 4 (BRD4), a known regulator of *MYC* transcription; its degradation by MZ1 therefore leads to downregulation of *MYC* mRNA and protein levels[49] (Supplementary Fig. 8). We therefore assessed the antiproliferative and apoptotic effects of MZ1 and MYC-RIBOTAC from 0.1 to 10 μM in HeLa cells. The two chimeras inhibited proliferation to a similar extent at the 10 μM dose. However, MZ1 is more potent at provoking cell death—1 μM MZ1 led to reductions in cell viability that were observed in cells treated with 10 μM of MYC-RIBOTAC (Extended Data Fig. 11h). Finally, MYC-RIBOTAC showed a similar effect on the reduction of *MYC* mRNA and protein levels as well as cellular phenotype in MDA-MB-231 breast cancer cells (Extended Data Fig. 11i). Similarly, the effect of MYC-RIBOTAC on *MYC* mRNA levels was ablated in MDA-MB-231 cells after knockdown of RNase L (Extended Data Fig. 11j). To verify these results, we studied the effect of MYC-RIBOTAC on a *MYC* IRES-luciferase reporter assay in transfected HEK293T cells. MYC-RIBOTAC reduced luciferase levels (~40%), yet had no effect on a control luciferase reporter lacking the *MYC* IRES (Fig. 5f). Furthermore, mutating the IRES in the luciferase reporter such that the 5′U**UC**G/3′A**CC**C binding site forms base pairs (5′U**UC**G/3′A**AG**C) rendered the RIBOTAC inactive at both the mRNA and protein levels (Extended Data Fig. 11k).

To evaluate the selectivity of MYC-RIBOTAC transcriptome-wide in HeLa cells, we performed global RNA-seq analysis. Among 21,027 transcripts detected, 84 (0.40%) were significantly affected, with 51 transcripts upregulated and 33 downregulated (Fig. 5g and Extended Data Fig. 11l). Notably, the most significantly downregulated transcript was *EGR1*, a known downstream target of MYC[50] (Fig. 5g). Similar results were obtained from the transcriptome-wide analysis of the *MYC* siRNA, which

significantly affected 90 (0.38%) transcripts among the 23,741 detected (Extended Data Fig. 11l). Cumulative distribution analysis of 837 well-validated MYC-target genes[51,52,53] demonstrated a significant decrease in these transcripts after treatment with MYC-RIBOTAC ($P < 0.001$) or the *MYC* siRNA ($P < 0.001$) (Extended Data Fig. 11l). By contrast, the same analysis of 63 downstream targets of HIF-1α, another transcription factor that recognizes a similar DNA sequence (ACGTG)[54] to MYC (CCACGTG)[55], showed no significant change by MYC-RIBOTAC (Extended Data Fig. 11l).

The selectivity of MYC-RIBOTAC across the transcriptome was mirrored in the proteome. Only 28 (1.0%) out of 2,769 detectable proteins were significantly affected (16 downregulated and 12 upregulated; Extended Data Fig. 11m). Notably, 4 out of the 16 downregulated proteins are direct downstream targets of *MYC*. Eight of the 16 downregulated proteins are associated with the ribosome pathway, one of the main functional pathways of MYC-target genes[53]. Cumulative distribution analysis of the proteome data showed results consistent with those on the transcriptome, where MYC-RIBOTAC treatment significantly downregulated the protein levels of downstream targets of *MYC*, but not those of HIF-1α (Fig. 5h and Extended Data Fig. 11m). Collectively, these results confirm that MYC-RIBOTAC is selective transcriptome- and proteome-wide.

To further evaluate MYC-RIBOTAC in other relevant tumour cell lines, the human Burkitt lymphoma cell lines Namalwa and Raji and the leukaemia cell line HL-60 were tested, in which MYC is overexpressed due to translocations or amplification. Notably, the RNase L expression level varies among the cell lines by 25-fold—it is highly expressed in HeLa cells and has low expression in Raji Burkitt lymphoma cells (Extended Data Fig. 11n). The activity of MYC-RIBOTAC tracked with RNase L expression: *MYC* mRNA levels were reduced by around 50% in HeLa cells (Fig. 5c), 25% in Namalwa cells (Fig. 5i), 10% in HL-60 cells and there was no reduction in Raji cells (Extended Data Fig. 11n). As expected, MYC-Ctr was inactive in Namalwa Burkitt lymphoma cells (Fig. 5i). By contrast, MYC-RIBOTAC reduced *MYC* protein levels by around 50% after treatment with 5 μM (Fig. 5j), induced cell cycle arrest and provoked apoptosis (Fig. 5k,l, Extended Data Fig. 11n and Supplementary Fig. 9) and reduced colony formation by

about 50% (Fig. 5m). Notably, MYC-Ctr, which cannot degrade *MYC* mRNA, had no effect in any of these assays (Fig. 5k–m).

Programming small-molecule RNA degraders

Here we have demonstrated that inactive, biologically inert small molecules engaging an RNA target can be programmed into bioactive degraders. Applying this strategy to a miRNA precursor demonstrated that the RIBOTAC degrader reduced the levels of the miRNA in multiple cell lines covering two diseases and in vivo. This degradation inhibited cellular phenotypes and colonization processes that are driven by this miRNA. As a further proof of concept, this strategy was extended to convert biologically silent interactions into bioactive ones for two additional oncogenic targets, *JUN* and *MYC* mRNAs. Their RIBOTAC degraders, derived from binders designed by Inforna, degraded the desired targets and disabled transcriptional and proteomic programs driven by these oncoproteins.

Collectively, this study provides examples of converting functionally silent binders into potent degraders that can selectively downregulate disease-causing RNAs. To our knowledge, all previously reported bioactive small molecules that target miRNAs bind in or nearby functional sites, which in this study constituted only around 30% of human miRNAs. Much less is known about functional sites in mRNAs, making the RIBOTAC strategy perhaps even more important for this class of targets. Complementarily, about 50% of a given RNA target is unstructured, providing ideal target sites for designer oligonucleotides. Small molecules provide a means to target the remaining structured regions, which is probably no longer limited to functional regions using the RIBOTAC approach.

Methods

General methods

All DNA templates and primers were purchased from Integrated DNA Technologies (IDT) and used without further purification. Chemically synthesized RNA oligonucleotides and oligonucleotide competitors were

obtained from Dharmacon and deprotected according to the manufacturer's recommended protocol. After deprotection, the RNAs were desalted using a PD-10 Sephadex column (GE Healthcare) according to the manufacturer's protocol. All oligonucleotide concentrations were determined by their absorption at 260 nm at 90 °C and the corresponding extinction coefficient provided by the manufacturer. The compound MZ1 was purchased from MedChemExpress.

All autoradiographical images and ethidium bromide staining images were obtained on the Typhoon FLA9500 variable mode imager (GE Healthcare). The Western blot images were obtained using the AFP Imaging Mini-Medical/90 system. These images were quantified using Image J (v.1.8.0_112). All numerical calculations were performed using Microsoft Excel (Office 365) and GraphPad 8 (v.8.0.2).

Statistics and reproducibility

Experimental results shown as representative blots or gel autoradiograms were successfully replicated two or more times to ensure the reproducibility of the reported findings. The number and type of replicates as well as statistical significance are provided in the figure legends.

In vitro Chem-CLIP binding site mapping

Pre-miR-155 RNA (500 pmol) was folded in 1× folding buffer (FB; 20 mM HEPES, pH 7.5, 150 mM NaCl, and 5 mM KCl) by heating at 95 °C for 30 s followed by cooling on ice for 5 min. Next, 50 nM of pre-miR-155-Chem-CLIP was added to the folded pre-miR-155, and the sample was incubated at 37 °C overnight. Cross-linked pre-miR-155 was captured on Dynabeads and eluted according to the manufacturer's protocol. Reverse transcription of the pulled-down RNA was performed using a gene-specific primer (5′-CAGACGTGCTCTTCCGATCTCTGTTAATGCTAATAT-3′) and Superscript III (SSIII) according to the manufacturer's protocol. The RT reaction was processed using RNAClean XP (Beckman Coulter) according to the manufacturer's protocol. A ssDNA adaptor (5′phosphate-NNNAGATCGGAAGAGCGTCGTGTAG-3C spacer) was ligated to the purified RT product using T4 RNA ligase I (New England BioLabs (NEB))

and then RNA Clean XP was added according to the manufacturer's protocol. The ligated cDNA was amplified using Phusion polymerase (NEB) using the following primers: 5′-CTACACGACGCTCTTCCGATCT-3′ and 5′-CAGACGTGCTCTTCCGAT-3′. The PCR reaction was purified with a denaturing 10% acrylamide gel and ethanol precipitation. The purified PCR product was cloned into NEB's pminiT 2.0 vector according to the manufacturer's protocol and sequenced by Eton Biosciences.

Tissue culture

MDA-MB-231 triple-negative breast cancer cells (HTB-26; ATCC) were cultured in RPMI medium with l-glutamine and 25 mM HEPES (Corning; 10-041) supplemented with 10% (v/v) fetal bovine serum (FBS; Sigma-Aldrich; F2442) and 1× penicillin–streptomycin solution (Gibco; 15140122).

MCF-10a breast epithelial cells (CRL-10317; ATCC) were cultured in DMEM/F12 50/50 medium with l-glutamine and 15 mM HEPES (Corning; 10-092-CV) containing 20% (v/v) FBS, 1× antibiotic–antimycotic solution (Corning; 30-004-CI), 20 ng ml^{-1} human epidermal growth factor (Pepro Tech; GMP100-15), 100 μg ml^{-1} insulin and 0.5 mg ml^{-1} hydrocortisone (Sigma-Aldrich; H0888).

CFPAC1 ductal adenocarcinoma epithelial cells (CRL-1918; ATCC) were cultured in 1× IMDM medium (Gibco, 12440053) with 10% (v/v) FBS and 1× antibiotic–antimycotic solution.

HUVECs (Lonza, CC-2517) were cultured in EGM prepared using the EGM-2 bullet kit (Lonza; CC-3162) according to the manufacturer's protocol.

HeLa cells (CCL-2, ATCC) and HEK293T cells (CRL-3216, ATCC) were cultured in 1× DMEM medium with 4.5 g l^{-1} glucose (Gibco, 11965092) supplemented with 2 mM glutamine or 1% Glutagro (Corning; 25-015-CI), 1× penicillin–streptomycin solution and 10% (v/v) FBS.

MIA PaCa-2 pancreatic adenocarcinoma cells (CRL-1420; ATCC) and CRISPR lines derived from MIA PaCa-2 pancreatic adenocarcinoma cells

(Supplementary Fig. 6) were cultured in 1× DMEM medium with 4.5 g l^{-1} glucose supplemented with 1% Glutagro, 1× penicillin–streptomycin solution and 10% (v/v) FBS.

Namalwa (CRL-1432) and Raji (CCL-86) Burkitt lymphoma cells, and HL-60 (CCL-240) myeloid leukaemia cells, were obtained from ATCC and were cultured in RPMI medium with l-glutamine and 25 mM HEPES supplemented with 10% (v/v) FBS and 1× penicillin–streptomycin solution.

The passage number for all cell lines was <20, except for HUVECs (<6) and HeLa cells (<30). All cells were checked to be free of mycoplasma contamination before conducting experiments using a PCR Mycoplasma Test Kit (PromoCell).

Forced expression of pre-miR-155 in MCF-10a cells

The plasmid to overexpress wild-type pre-miR-155 were purchased from Genecopoeia (HmiR0358) and the mutant plasmids were custom synthesized by GeneScript. MCF-10a cells were seeded at around 70% confluency in 60 mm dishes and transfected with a plasmid (2 µg per dish) encoding either wild-type or mutant pre-miR-155 (Extended Data Fig. 9a) using Lipofectamine 3000 (Invitrogen) according to the manufacture's protocol. The cells were incubated in the transfection cocktail for 6 h followed by changing to fresh growth medium and incubated for an additional 16 h. Cells were then trypsinized and seeded for biological experiments as described in the following sections.

Cellular Chem-CLIP and competitive Chem-CLIP analysis of pre-miR-155, MYC and JUN

General protocol for cellular Chem-CLIP

After compound treatment as specified for each target below, cells were washed once with 1× DPBS and irradiated by ultraviolet light for 15 min. Total RNA was extracted using the Zymo Quick-RNA MiniPrep Kit according to the manufacturer's protocol. The input RNA (10 µg) was incubated with azide-modified agarose beads (Click Chemistry Tools)

according to the manufacturer's protocol, followed by adding a total of 45 μl reaction solution with 1:1:1 of ascorbic acid (250 mM), $CuSO_4$ (10 mM), THPTA (tris-hydroxypropyltriazolylmethylamine; 50 mM). After incubating at 37 °C for 2 h, the beads were washed six times with washing buffer (10 mM Tris-HCl, pH 7, 4 M NaCl, 1 mM EDTA and 0.1% Tween-20) and resuspended in 100 μl release solution containing 200 mM TCEP and 400 mM K_2CO_3. The mixture was incubated at 37 °C for 30 min, followed by adding 50 μl of 800 mM iodoacetamide and incubating for another 30 min. The supernatant containing RNA was carefully transferred to a clean tube and purified by RNAClean XP beads (Beckman Coulter) according to the manufacturer's protocol. The purified RNA was analysed using RT–qPCR as described in the 'RT–qPCR analysis of mRNAs and pri-, pre- and mature miRNA levels' section.

Pre-miR-155 Chem-CLIP

MDA-MB-231 cells, grown to about 80% confluency in 100-mm-diameter dishes, were treated with DMSO (vehicle; 0.1% (v/v), the final concentration in all compound-treated cells), 100 nM pre-miR-155-Chem-CLIP or 100 nM control probe Ac-Chem-CLIP for 6 h. Competitive Chem-CLIP was performed by pretreating MDA-MB-231 cells with pre-miR-155-amide (0–1 μM as indicated) for 2 h followed by addition of pre-miR-155-Chem-CLIP (100 nM) and incubating the cells for 16 h.

JUN Chem-CLIP

MIA PaCa-2 cells, grown to about 90% confluency in 100-mm-diameter dishes, were treated with DMSO (vehicle), JUN-Chem-CLIP or Ctrl-Chem-CLIP at the indicated concentrations for 6 h. Competitive Chem-CLIP was performed by pretreating MIA PaCa-2 cells with JUN-binder at the indicated concentrations for 2 h followed by the addition of JUN-Chem-CLIP and incubating cells for another 5 h. After cross-linking and RNA isolation as described in the 'General protocol for cellular Chem-CLIP', the RNA samples were supplemented with a click reaction mixture consisting of disulfide biotin azide (Click Chemistry Tools, 1168, 0.5 μl, 100 mM), $CuSO_4$ (1 μl, 10 mM), THPTA (1 μl, 50 mM) and sodium ascorbate (1 μl,

250 mM). The reaction was incubated at 37 °C for 3 h. After the incubation period, 100 µl of Streptavidin beads (Dynabeads MyOne Streptavidin C1 beads; Thermo Fisher Scientific) was added, and the samples were incubated at room temperature for 1 h. The washing and elution steps were identical to those described in the 'General protocol for cellular Chem-CLIP' section.

MYC Chem-CLIP

HeLa cells, grown to about 80% confluency in 60-mm-diameter dishes, were treated with vehicle, MYC-Chem-CLIP or control probe Ctrl-Chem-CLIP for 5 h. MYC competitive Chem-CLIP was performed by pretreating HeLa cells with MYC-binder at the indicated concentration for 2 h followed by addition of MYC-Chem-CLIP (10 µM) and incubating the cells for 5 h. The remaining steps were performed as described in the 'General protocol for cellular Chem-CLIP' section.

RT–qPCR analysis

General Protocol for RT–qPCR analysis

Total RNA was extracted using the Zymo Quick-RNA MiniPrep Kit according to the manufacturer's protocol. For pri- and pre- miRNAs and mRNAs, RT was performed using 200 ng of RNA with a Qscript cDNA Synthesis Kit (QuantaBio). For mature miRNAs, the RT reaction was performed using 200 ng of RNA with the miScript II RT Kit in a total volume of 20 µl (Qiagen). Subsequent qPCR analysis (see Supplementary Table 2 for primer sequences and Supplementary Figs. 2–7 for primer validation) using Power SYBR Green Master Mix (Life Technologies) or a TaqMan assay and a Applied Biosystems QS5 384-well PCR system (software v.1.3.0). For mature miR-155 levels, TaqMan assays were performed using the ipu-miR-155 Taqman Assay (Thermo Fisher Scientific, 467534). Data were analysed using the $\Delta\Delta C_t$ method as described previously[13].

RT–qPCR analysis of mRNA and pri-, pre- and mature miRNA levels

MDA-MB-231 cells or MCF-10a cells were seeded in 12-well plates at around 60% confluency. Cells were treated with compound, LNA-155 miRCURY LNA Power Inhibitor (Qiagen; 5′UUAAUGCUAAUCGUGAUAGGGGU) or vehicle (0.1% (v/v) DMSO, the final concentration in all compound-treated samples) at the indicated concentrations for 48 h in growth medium (unless noted otherwise; Extended Data Fig. 6e). For washout experiments, MDA-MB-231 cells (around 40% confluency) were treated with pre-miR-155-RIBOTAC at the indicated concentration in growth medium for 48 h. After the treatment period, the cells were washed with 1× DPBS and fresh growth medium without the compound was added. Total RNA was then collected from the cells at the indicated timepoints after treatment (12, 24 and 36 h). For competitive experiments with the binder and the RIBOTAC, both compounds were added to the growth medium at the indicated concentrations, applied to the cells and incubated for 48 h. For miRNA profiling, cells were treated with pre-miR-155-RIBOTAC (100 nM) or **LNA-155** (50 nM) in growth medium for 48 h.

CFPAC cells were seeded in 12-well plates at about 50% confluency and treated with compounds at the indicated concentrations for 48 h. HUVECs cells were seeded in 12-well plates at around 60% confluency and treated with compound prepared in growth medium at the indicated concentrations for 48 h. The competitive experiment with the binder and the RIBOTAC was performed by adding both compounds to the growth medium to the desired concentrations, applying the mixture to the cells and incubating for 48 h.

RT–qPCR analysis of *JUN* mRNA

MIA PaCa-2 cells, whether unmodified or CRISPR-modified cell lines, were grown to around 40% confluency in 12-well plates and then treated with vehicle (DMSO; 0.1% (v/v), the final concentration in all compound-treated samples) or compound in growth medium at the indicated concentration. The siRNA directed at *JUN* (Dharmacon; L-003268-00-0005) or a scrambled control siRNA (Dharmacon; D-001810-10-05) was transfected with Lipofectamine 3000 (Invitrogen, L3000001) according to the manufacturer’s instructions. After RT as described above, qPCR amplification was performed using the TaqMan gene expression assay

(*GAPDH*: Hs03929097_g1, 4331182; *JUN*: Hs00277190_s1, 4331182) and the TaqMan Fast Advanced MasterMix. The same protocol was used for MIA PaCa-2 cells in which RNase L was knocked down by CRISPR and cells for which a control guide RNA was used.

RT–qPCR analysis of *MYC* mRNA

HeLa cells and MDA-MB-231 cells (whether unmodified or CRISPR-modified cell lines), grown to around 30% or 50% confluency in 12-well plates, respectively, were treated with vehicle (DMSO; 0.1% (v/v), the final concentration in all compound-treated samples) or compound of interest in growth medium for 24 h. After 24 h, the medium was removed, and the cells were treated with fresh medium containing compound for additional 24 h. For *MYC* siRNA studies, cells were seeded in 12-well plates and 1 nM of *MYC* siRNA (SantaCruz, sc-29226) or scrambled control (SantaCruz, sc-37007) was transfected with Lipofectamine 3000 according to the manufacturer's recommended protocol. Namalwa (0.4×10^6 cells per ml, 2 ml), HL-60 (0.5×10^6 cells per ml, 2 ml) and Raji (0.7×10^6 cells per ml, 2 ml) cells were treated with vehicle, MYC-RIBOTAC or MYC-Ctr for 48 h. For Namalwa, HL-60 and Raji cells, RNase L mRNA levels were also measured using RT–qPCR. For studies performed in MCF-10a and MDA-MB-231 cells, TaqMan assays were also performed using the TaqMan gene expression assay (*GAPDH*: Hs03929097_g1, 4331182; *MYC*: Hs00153408_m1, 4331182) and the TaqMan Fast Advanced MasterMix.

RT–qPCR analysis of *MYC* IRES-luciferase mRNA

HEK293T cells were transfected and treated as described in the 'MYC luciferase reporter assay' section except in 12-well instead of 96-well plates. After 48 h, the total RNA was extracted and RT–qPCR was performed as described for mRNAs in the 'General protocol for RT–qPCR analysis' section.

Western blot analysis

General protocol for western blotting

After treatment as specified for each target below, total protein was collected using M-PER Extraction Reagent (Thermo Fisher Scientific), and protein concentration was measured using the Pierce Micro BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer’s protocol. Approximately 20 µg of total protein, except for analysis of SOCS1 protein (which used 40 µg due to its low abundance), was resolved on a 10% SDS–polyacrylamide gel. After transferring to a PVDF membrane and blocking the membrane with 1× tris-buffered saline supplemented with 0.1% (v/v) Tween-20 (TBST; 50 mM Tris-Cl, pH 7.5, 150 mM NaCl and 0.1% (v/v) Tween-20) with 5% (w/v) milk, the membrane was incubated with the primary antibody (see below for each target) in 1× TBST with 5% (w/v) milk at 4 °C for 16 h. The membrane was washed three times with 1× TBST (5 min per wash) at room temperature. The blot was then incubated with IgG horseradish-peroxidase (HRP) secondary antibody conjugate in 1× TBST with 5% milk at room temperature for 1 h. After washing five times with 1× TBST (5 min per wash), the target protein was detected by using SuperSignal West Pico Chemiluminescent Substrate (Pierce Biotechnology) and quantified using ImageJ. The membrane was then stripped (200 mM glycine, pH 2.2, 4 mM SDS, 1% (v/v) Tween-20) at room temperature for 30 min and blotted again with either β-actin or GAPDH for normalization.

Western blot analysis of SOCS1

MDA-MB-231 cells were seeded in six-well plates at about 60% confluency and treated with pre-miR-155-RIBOTAC at the indicated concentrations for 48 h. The anti-SOCS1 primary antibody (Cell Signaling Technology (CST), 3950S) was used at 1:2,500 dilution followed by anti-rabbit IgG-HRP secondary antibody conjugate (CST, 7074S) at 1:5,000 dilution. After stripping, β-actin primary antibody (CST, 3700S) was used at 1:5,000 dilution, followed by anti-mouse IgG HRP secondary antibody conjugate (CST, 7076S) at 1:10,000 dilution.

Western blot analysis of VHL

HUVECs (around 60% confluency) were seeded in six-well plates and treated with pre-miR-155-RIBOTAC at the indicated concentration for 48 h. The anti-VHL primary antibody (CST, 68547S) was used at 1:1,000 dilution

followed by anti-rabbit IgG HRP (CST, 7074S) at 1:5,000 dilution. β-Actin was detected as described above in the 'Western blot analysis of SOCS1' section.

Western blot analysis of JUN and RNase L

MIA PaCa-2 cells were seeded in six-well plates and treated with vehicle (0.1% (v/v) DMSO; the final concentration in all compound-treated samples) or compound at the indicated concentrations in growth medium for 72 h once they reached around 40% confluency. The anti-JUN primary antibody (CST, 9165S) or anti-RNase L antibody (CST, D4B4J) was used at 1:1,000 dilution followed by anti-rabbit IgG HRP (CST, 7074S) at 1:10,000 dilution. β-Actin was detected as described above in the 'Western blot analysis of SOCS1' section.

Western blot analysis of MYC

HeLa cells in six-well plates (around 30% confluency) or MDA-MB-231 cells (about 60% confluency) were seeded in six-well plates and treated with vehicle or compound at the indicated concentrations in growth medium for 24 h. The compound-containing growth medium was removed and replaced with fresh growth medium containing compound. The cells were incubated for an additional 24 h at which point total protein was collected as described above. Namalwa cells (0.4×10^6 cells per ml, 10 ml) were treated with vehicle, MYC-RIBOTAC or MYC-Ctr for 48 h (single dose), and then total protein was collected. The anti-MYC antibody (CST, 5605S) was used at 1:1,000 dilution followed by anti-rabbit IgG HRP (CST, 7074S) at 1:5,000 dilution. The anti-GAPDH primary antibody (CST, 51332S) was used at 1:3,000 dilution followed by direct imaging (no secondary antibody needed as this antibody is an HRP conjugate). The membrane was not stripped and was directly blotted with anti-GAPDH antibodies for normalization.

Western blot analysis of BRD4

The same procedure was followed as described in the 'Western blot analysis of MYC' section except that the primary antibody was anti-BRD4 (Cell

Signaling Technologies, 13440S) at 1:1,000 dilution. The effect of MZ1 on BRD4 and MYC protein levels was validated and shown in Supplementary Fig. 8.

SOCS1 luciferase reporter assay (pre-miR-155)

HEK293T cells in 60-mm-diameter dishes (around 80% confluency) were co-transfected with 4 µg of a plasmid encoding SOCS1 fused to luciferase (Genecopoeia, HmiT021399-MT06) and 1 µg of the plasmid expressing wild-type pre-miR-155 using jetPRIME (Polypuls, 101000027) according to the manufacturer's recommended protocol. The cells were seeded to 96-well plates (7,000 cells per well) after transfection. After incubating for 12 h, the cells were treated with DMSO (vehicle, 0.1% (v/v)) or pre-miR-155-RIBOTAC. A dose–response was measured by treating pre-miR-155-RIBOTAC at 0, 10 and 100 nM for 48 h and a time course was measured by treating pre-miR-155-RIBOTAC (100 nM) for 6, 24, 48 and 72 h. The Dual-Glo Luciferase Assay System (Promega, E2920) was used to measure luciferase activity, according to the manufacturer's instructions.

MYC luciferase reporter assay

The *MYC* IRES luciferase plasmid was custom-synthesized by GenScript (U6617GC310) by inserting the *MYC* IRES sequence between the two Click Beetle luciferases (*CBR* and *CBG*) within the pcDNA5 vector. HEK293T cells in 60-mm-diameter dishes (around 80% confluency) were co-transfected with 2 µg of a plasmid and 500 ng of a plasmid expressing Renilla luciferase (normalization) using jetPRIME (Polyplus, 101000027) according to the manufacturer's recommended protocol. Control experiments were performed by co-transfecting 2 µg of a plasmid encoding luciferase lacking the IRES (GenScript) and 500 ng of a plasmid expressing *Renilla* luciferase (for normalization). The cells were trypsinized and seeded into 96-well plates (7,000 cells per well). After incubating for 12 h, the cells were treated with vehicle or MYC-RIBOTAC (10 µM) for the indicated time. Click Beetle (which uses the same substrate as firefly luciferase) and *Renilla* luciferase activities were measured using the Dual-Glo Luciferase Assay System (Promega, E2920) according to the manufacturer's instructions.

To construct a mutant *MYC* IRES plasmid in which the small-molecule-binding motif was mutated into base pairs, the Q5 Site-Directed Mutagenesis Kit (NEB, E0554S) was used according to the manufacturer's recommendations. The primers used to create the mutant were: 5′CGCCTCTGGC**GA**AGCCCTCCCG and 5′AAGCCCCCTATTCGCTCC (IDT), where the bold nucleotide indicates the point mutation. The mutagenesis product was transformed into chemically competent *Escherichia coli* (included in the kit) and selected using ampicillin-agar plates. The mutant plasmid sequence was confirmed by Sanger sequencing, and the plasmid was tested in HeLa cells according to the same procedure as described above.

Methylcellulose colony assays

Namalwa Burkitt lymphoma cells (1,000 cells per ml, 3 ml) were treated with MYC-Ctr (10 μM) or MYC-RIBOTAC (10 μM) for 48 h (single dose). The cells were then collected by centrifugation for 5 min at 1,000 rpm, and 1×10^3 viable cells were resuspended in 400 μl of IMDM medium containing 2% (v/v) FBS. The cells were then added to 4 ml of Methocult (STEMCELL Technology, H4230), and 1.1 ml was carefully spread into a SmartDish (STEMCELL Technology, 27370). Cells were incubated for 7–10 days, and colonies were counted manually.

Cell cycle analysis

Namalwa (0.4×10^6 cells per ml, 2 ml), HL-60 (0.5×10^6 cells per ml, 2 ml) and Raji (0.7×10^6 cells per ml, 2 ml) cells were treated with vehicle, MYC-RIBOTAC (10 μM) or MYC-Ctr (10 μM) for 48 h (single dose). To assess the percentage of cells at each stage of the cell cycle, 1×10^6 cells were fixed dropwise with 70% (v/v) ethanol at room temperature, and then stored for 24 h at −20 °C. Fixed cells were washed twice with 1× PBS, and then treated with 50 μl of 100 μg ml^{-1} RNase A (Sigma-Aldrich, R-6148) for 5 min room temperature. The cells were then stained with 200 μl of 50 μg ml^{-1} propidium iodide; Sigma-Aldrich, P4170) prepared in 1× PBS. The stained cells were analysed for the percentage in the G1, S and G2 phases of the cell cycle in a BD LSRII (BD Biosciences) flow cytometer. Cell doublets were excluded from analysis.

Drug metabolism and pharmacokinetic studies (pre-miR-155)

Female C57Bl/6 mice (*n* = 3) were treated once with either pre-miR-155-RIBOTAC or pre-miR-155-amide-binder at 1 mg per kg (10:10:80 of DMSO:Tween-80:H_2O) by i.p. injection. The plasma was then collected at 0, 15, 30, 60, 120, 240, 360, 480 and 1,440 min to determine the concentration in plasma.

In vivo studies to assess lung colonization of breast cancer cells (pre-miR-155)

Mice were housed in individually ventilated (IVC), JAG 75 cages with micro-isolator lids. HEPA filtered air was supplied into each cage at a rate of 60 air exchanges per h. The dark–light cycle was set for 08:00 (on)–20:00 (off). The temperature was maintained at 72 ± 2 °F (22.2 ± 1.1 °C). Humidity was maintained at 30–70%. Female NOD/SCID mice ($n = 8$ per group; 5–7 weeks; The Jackson Laboratory) were used for breast cancer studies as described previously[56], and all studies were approved by The Scripps Research Institute IACUC (protocol no. 16-025). The maximal tumour size allowable under this protocol is 500 mm^3 or 2 cm in diameter.

In brief, mice were intravenously injected (tail vein) with MDA-MB-231-Luc cells (0.8×10^6 cells per mouse). Compound treatment began 5 days later as determined by luciferase activity, where the averaged luciferase signal in the lung tissue is ten times higher than background measured from neighbouring tissues. Luciferase activity was measured by i.v. injection of luciferin (150 mg per kg) every other day as measured by LagoX (Spectral Instruments). Mice were split into two groups: vehicle group (10:10:80 of DMSO:Tween-80:H_2O) or 1 mg per kg of either pre-miR-155-RIBOTAC or pre-miR-155-amide in the same formulation. Compound or vehicle was delivered by i.p. injection every other day.

After 30 days, mice were euthanized by affixiation with CO_2 in accordance with guidelines provided by the American Veterinarian Medical Association. For each group, the lungs from five mice were perfused with 1× PBS, collected and immediately fixed in Bouin's solution (Sigma-Aldrich, HT10132-1L) to image nodules. After imaging, the lungs were sectioned and stained with H&E staining or analysed using FISH to assess pre-miR-155 levels. The FISH probe and qPCR primers are specific to human mature and pre-miR-155; mouse pre-miR-155 has a different secondary structure lacking the binding site for pre-miR-155-RIBOTAC. Human and mouse SOCS1 protein cannot be distinguished by available antibodies; therefore, SOCS1 protein levels were not determined. For the

remaining three mice from each group, the lungs were collected and snap-frozen for extracting total RNA to measure mature-miR-155 levels by RT–qPCR.

To image lung nodules, lungs were fixed in 60 ml of Bouin's solution for 24 h followed by briefly rinsing in 10% neutral buffered formalin solution (Harleco). After counting nodules, the lungs were washed with 60 ml of 10% (v/v) formalin solution four times (8 h per wash with gentle shaking at room temperature). Paraffin-embedded sections were generated by the Histology Core at The Scripps Research Institute using a Sakura Tissue-Tek VIP5 paraffin processor. H&E staining was performed on the Leica ST5010 Auto-Stainer XL at The Scripps Research Institute's Histology Core.

For FISH analysis to assess mature miR-155 levels, the sections were incubated at 60 °C for 15 min followed by deparaffination with xylene. The sections were rinsed once in each of the following solutions: 95%, 80%, 50% and 0% ethanol in water. The sections were then treated with 20 μg ml^{-1} proteinase K in 50 mM Tris, pH 7.5, for 10–20 min at 37 °C, followed by rinsing in water five times. The sections were then treated with ice-cold 20% (v/v) acetic acid for 20 s to permeabilize cells, followed by rinsing with Nanopure water three times. The sections were dehydrated by rinsing in 100% ethanol. The sections were then prepared for FISH analysis by rinsing the sections in 40% formamide/2× SSC (0.3 M sodium chloride and 0.03 M sodium citrate, pH 7.0) buffer for 5 min. The slides were then incubated with 40% formamide/2× SSC buffer containing 0.2 μM FITC-labelled DNA probe (5′-AACCCCTATCACGATTAGCATTAA), 2 μg ml^{-1} BSA, 10 nM tRNA (Roche, 10109517001) in a humidified chamber at 37 °C for 16 h. The slides were then rinsed three times with 2× SSC, and once with 1× DPBS, followed by addition of an anti-FITC antibody (Abcam, 6656; 1:500 dilution in 1×TBST buffer containing 1% (v/v) goat serum (Sigma-Aldrich)). After incubation at room temperature for 2 h, the slides were rinsed three times with 1× DPBS and visualized using a DAB Substrate Kit (Abcam, 64238) according to the manufacturer's protocol. Images were acquired using a bright-field microscope (Leica DM800) and quantified using ImageJ.

The lungs from the other three mice were frozen at −80 °C and homogenized using tissue plastic homogenizing probes (Omni International, 10062-782). The RNA extraction and RT–qPCR analysis were performed the same as described above in the 'RT–qPCR analysis of mRNAs and pri-, pre- and mature miRNA levels' section.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data and materials are available from the corresponding author on reasonable request. The data for global proteomics and RNA-seq studies are available at Mendeley Data (https://doi.org/10.17632/xgr83xy8pm.1). The sequencing data for plasmids used in this study are available at Mendeley Data (https://doi.org/10.17632/9zgvv67j7s.1). The raw sequencing data have been deposited at the Sequence Read Archive under BioProject ID PRJNA914317. The R-Bind database[17] is publicly accessible (https://rbind.chem.duke.edu). The Inforna database is available online (https://rnainforna.com). Access is freely granted for academic users after completion of a software license agreement (https://disney.scripps.ufl.edu/wp-content/uploads/2020/05/software-licensing-agreement-1.pdf). Source data are provided with this paper.

Code availability

No unique code was used in the described data analyses.

References

1. Costales, M. G. et al. Small-molecule targeted recruitment of a nuclease to cleave an oncogenic RNA in a mouse model of metastatic cancer. *Proc. Natl Acad. Sci. USA* **117**, 2406–2411 (2020).

2. Cooper, T. A., Wan, L. & Dreyfuss, G. RNA and disease. *Cell* **136**, 777–793 (2009).

3. Frazier, K. S. Antisense oligonucleotide therapies: the promise and the challenges from a toxicologic pathologist's perspective. *Toxicol. Pathol.* **43**, 78–89 (2015).

4. Lenartowicz, E. et al. Antisense oligonucleotides targeting influenza A segment 8 henomic RNA inhibit viral replication. *Nucleic Acid Ther.* **26**, 277–285 (2016).

5. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.* **5**, 451–463 (2004).

6. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).

7. Rizvi, N. F. et al. Targeting RNA with small molecules: identification of selective, RNA-binding small molecules occupying drug-like chemical space. *SLAS Discov.* **25**, 384–396 (2020).

8. Costales, M. G., Childs-Disney, J. L., Haniff, H. S. & Disney, M. D. How we think about targeting RNA with small molecules. *J. Med. Chem.* **63**, 8880–8900 (2020).

9. Grigalunas, M., Brakmann, S. & Waldmann, H. Chemical evolution of natural product structure. *J. Am. Chem. Soc.* **144**, 3314–3329 (2022).

10. Wicks, S. L. & Hargrove, A. E. Fluorescent indicator displacement assays to identify and characterize small molecule interactions with RNA. *Methods* **167**, 3–14 (2019).

11. Velagapudi, S. P. et al. Approved anti-cancer drugs target oncogenic non-coding RNAs. *Cell Chem. Biol.* **25**, 1086–1094 (2018).

12. Disney, M. D. Targeting RNA with small molecules to capture opportunities at the intersection of chemistry, biology, and medicine. *J.*

Am. Chem. Soc. **141**, 6776–6790 (2019).

13. Velagapudi, S. P. et al. Defining RNA-small molecule affinity landscapes enables design of a small molecule inhibitor of an oncogenic concoding RNA. *ACS Cent. Sci.* **3**, 205–216 (2017).

14. Bajusz, D., Rácz, A. & Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7**, 20 (2015).

15. Haniff, H. S. et al. Design of a small molecule that stimulates vascular endothelial growth factor A enabled by screening RNA fold–small molecule interactions. *Nat. Chem.* **12**, 952–961 (2020).

16. Disney, M. D. et al. Inforna 2.0: a platform for the sequence-based design of small molecules targeting structured RNAs. *ACS Chem. Biol.* **11**, 1720–1728 (2016).

17. Morgan, B. S., Forte, J. E., Culver, R. N., Zhang, Y. & Hargrove, A. E. Discovery of key physicochemical, structural, and spatial properties of RNA-targeted bioactive ligands. *Angew. Chem. Int. Ed. Engl.* **56**, 13498–13502 (2017).

18. Liu, B. et al. Analysis of secondary structural elements in human microRNA hairpin precursors. *BMC Bioinform.* **17**, 112 (2016).

19. Han, Y. et al. Structure of human RNase L reveals the basis for regulated RNA decay in the IFN response. *Science* **343**, 1244–1248 (2014).

20. Wreschner, D. H., McCauley, J. W., Skehel, J. J. & Kerr, I. M. Interferon action—sequence specificity of the ppp(A2′p)$_n$A-dependent ribonuclease. *Nature* **289**, 414–417 (1981).

21. Floyd-Smith, G., Slattery, E. & Lengyel, P. Interferon action: RNA cleavage pattern of a (2′-5′)oligoadenylate-dependent endonuclease. *Science* **212**, 1030–1032 (1981).

22. Le Roy, F. et al. The 2–5A/RNase L/RNase L inhibitor (RNI) pathway regulates mitochondrial mRNAs stability in interferon α-treated H9 cells. *J. Biol. Chem.* **276**, 48473–48482 (2001).

23. Rühling, A., Galla, H. J. & Glorius, F. A remarkably simple hybrid surfactant-NHC ligand, its gold-complex, and application in micellar catalysis. *Chem. Eur. J.* **21**, 12291–12294 (2015).

24. Wang, D. et al. A remarkably simple class of imidazolium-based lipids and their biological properties. *Chem. Eur. J.* **21**, 15123–15126 (2015).

25. Guedes, J. R. et al. Early miR-155 upregulation contributes to neuroinflammation in Alzheimer's disease triple transgenic mouse model. *Hum. Mol. Genet.* **23**, 6286–6301 (2014).

26. Jiang, S. et al. MicroRNA-155 functions as an OncomiR in breast cancer by targeting the suppressor of cytokine signaling 1 gene. *Cancer Res.* **70**, 3119–3127 (2010).

27. Guan, L. & Disney, M. D. Covalent small-molecule-RNA complex formation enables cellular profiling of small-molecule-RNA interactions. *Angew. Chem. Int. Ed. Engl.* **52**, 10010–10013 (2013).

28. Zheng, R. et al. Long non-coding RNA XIST inhibited breast cancer cell growth, migration, and invasion via miR-155/CDX1 axis. *Biochem. Biophys. Res. Commun.* **498**, 1002–1008 (2018).

29. Velagapudi, S. P., Li, Y. & Disney, M. D. A cross-linking approach to map small molecule-RNA binding sites in cells. *Bioorg. Med. Chem. Lett.* **29**, 1532–1536 (2019).

30. Békés, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. *Nat. Rev. Drug Discov.* **21**, 181–200 (2022).

31. Farnaby, W. et al. BAF complex vulnerabilities in cancer demonstrated via structure-based PROTAC design. *Nat. Chem. Biol.* **15**, 672–680 (2019).

32. Bassi, Z. I. et al. Modulating PCAF/GCN5 immune cell function through a PROTAC approach. *ACS Chem. Biol.* **13**, 2862–2867 (2018).

33. Gantier, M. P. et al. Analysis of microRNA turnover in mammalian cells following Dicer1 ablation. *Nucleic Acids Res.* **39**, 5692–5703 (2011).

34. Tang, S. et al. Sweating the small stuff: microRNAs and genetic changes define pancreatic cancer. *Pancreas* **42**, 740–759 (2013).

35. Agarwal, V., Bell, G. W., Nam, J. W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015).

36. Johansson, J. et al. MiR-155-mediated loss of C/EBPβ shifts the TGF-β response from growth inhibition to epithelial-mesenchymal transition, invasion and metastasis in breast cancer. *Oncogene* **32**, 5614–5624 (2013).

37. Kong, W. et al. Upregulation of miRNA-155 promotes tumour angiogenesis by targeting VHL and is associated with poor prognosis and triple-negative breast cancer. *Oncogene* **33**, 679–689 (2014).

38. Mattiske, S., Suetani, R. J., Neilsen, P. M. & Callen, D. F. The oncogenic role of miR-155 in breast cancer. *Cancer Epidemiol. Biomarkers Prev.* **21**, 1236–1243 (2012).

39. Nanbru, C. et al. Alternative translation of the proto-oncogene c-*myc* by an internal ribosome entry site. *J. Biol. Chem.* **272**, 32061–32066 (1997).

40. Stoneley, M. et al. c-Myc protein synthesis is initiated from the internal ribosome entry segment during apoptosis. *Mol. Cell. Biol.* **20**, 1162–1169 (2000).

41. Lee, A. S., Kranzusch, P. J. & Cate, J. H. eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* **522**, 111–114 (2015).

42. Lee, A. S., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* **536**, 96–99 (2016).

43. Andrews, R. A.-O., Roche, J. & Moss, W. A.-O. ScanFold: an approach for genome-wide discovery of local RNA structural elements-applications to Zika virus and HIV. *PeerJ* **6**, e6136 (2018).

44. Vleugel, M. M., Greijer, A. E., Bos, R., van der Wall, E. & van Diest, P. J. c-Jun activation is associated with proliferation and angiogenesis in invasive breast cancer. *Hum. Pathol.* **37**, 668–674 (2006).

45. Zhang, Y. et al. Critical role of c-Jun overexpression in liver metastasis of human breast cancer xenograft model. *BMC Cancer* **7**, 145 (2007).

46. Ferrara, C. et al. Ki-67 and c-jun expression in pancreatic cancer: a prognostic marker? *Oncol. Rep.* **6**, 1117–1122 (1999).

47. Pelengaris, S., Khan, M. & Evan, G. c-MYC: more than just a matter of life and death. *Nat. Rev. Cancer* **2**, 764–776 (2002).

48. Stine, Z. E., Walton, Z. E., Altman, B. J., Hsieh, A. L. & Dang, C. V. MYC, metabolism, and cancer. *Cancer Discov.* **5**, 1024–1039 (2015).

49. Zengerle, M., Chan, K. H. & Ciulli, A. Selective small molecule induced degradation of the BET bromodomain protein BRD4. *ACS Chem. Biol.* **10**, 1770–1777 (2015).

50. Boone David, N., Qi, Y., Li, Z. & Hann Stephen, R. Egr1 mediates p53-independent c-Myc–induced apoptosis via a noncanonical ARF-dependent transcriptional mechanism. *Proc. Natl Acad. Sci. USA* **108**, 632–637 (2011).

51. Fernandez, P. C. et al. Genomic targets of the human c-Myc protein. *Genes Dev.* **17**, 1115–1129 (2003).

52. Dang, C. V. et al. The c-Myc target gene network. *Semin. Cancer Biol.* **16**, 253–264 (2006).

53. Kim, J., Lee, J.-H. & Iyer, V. R. Global identification of Myc target genes reveals its direct role in mitochondrial biogenesis and its E-box usage in vivo. *PLoS ONE* **3**, e1798 (2008).

54. Mimura, I. et al. Dynamic change of chromatin conformation in response to hypoxia enhances the expression of GLUT3 (SLC2A3) by cooperative interaction of hypoxia-inducible factor 1 and KDM3A. *Mol. Cell. Biol.* **32**, 3018–3032 (2012).

55. Seitz, V. et al. Deep sequencing of MYC DNA-binding sites in Burkitt lymphoma. *PLoS ONE* **6**, e26837 (2011).

56. Zhang, P. et al. Reprogramming of protein-targeted small-molecule medicines to RNA by ribonuclease recruitment. *J. Am. Chem. Soc.* **143**, 13044–13055 (2021).

57. Tran, T. & Disney, M. D. Identifying the preferred RNA motifs and chemotypes that interact by probing millions of combinations. *Nat. Commun.* **3**, 1125 (2012).

58. Rzuczek, S. G., Southern, M. R. & Disney, M. D. Studying a drug-like, RNA-focused small molecule library identifies compounds that inhibit RNA toxicity in myotonic dystrophy. *ACS Chem. Biol.* **10**, 2706–2715 (2015).

Acknowledgements

This work was supported by the National Institutes of Health grants R01 CA249180 (to M.D.D. and J.L.C.) and R35 NS116846-01 (to M.D.D.); the Cortner-Couch Endowed Chair for Cancer Research from the University of South Florida School of Medicine (to J.L.C.); the DFG (Deutsche Forschungsgesellschaft, ChemBIon, GRK2515; to F.G.); a Myotonic US Fellowship Research Grant (to R.I.B.); a National Ataxia Foundation Fellowship Research Grant (to R.I.B.); NCI Comprehensive Cancer Center Grant P30 CA076292, and by support from the State of Florida to the H. Lee Moffitt Cancer Center & Research Institute. We thank R. Hesterberg and the staff at the Flow Cytometry Core at Moffitt Cancer Center &

Research Institute for assisting in the flow cytometry experiments; and the staff at the Drug Metabolism and Pharmacokinetics Core and the Genomics Core at The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology for assisting in in vivo compound exposure and RNA-seq studies, respectively.

Author information

Author notes

1. These authors contributed equally: Yuquan Tong, Yeongju Lee, Xiaohui Liu, Jessica L. Childs-Disney

Authors and Affiliations

1. Department of Chemistry, The Scripps Research Institute & The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, Jupiter, FL, USA

   Yuquan Tong, Yeongju Lee, Xiaohui Liu, Jessica L. Childs-Disney, Blessy M. Suresh, Raphael I. Benhamou, Matthew G. Costales, Hafeez S. Haniff, Daniel Abegg, Elizabeth Lekah, Maison Grefe, Tenghui Wang, Quentin M. R. Gibaut, Alexander Adibekian & Matthew D. Disney

2. Department of Tumor Biology, Moffitt Cancer Center & Research Institute, Tampa, FL, USA

   Chunying Yang, Weimin Li & John L. Cleveland

3. Max Planck Institute of Molecular Physiology, Dortmund, Germany

   Sonja Sievers & Herbert Waldmann

4. Compound Management and Screening Center, Dortmund, Germany

   Sonja Sievers & Herbert Waldmann

5. Organisch-Chemisches Institut, University of Münster, Münster, Germany

Tristan Wegner, Tiffany O. Paulisch & Frank Glorius

6. Bioinformatics and Statistics Core, The Scripps Research Institute and The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, Jupiter, FL, USA

Gogce Crynen

7. Histology Core, The Scripps Research Institute and The Herbert Wertheim UF Scripps Institute for Biomedical Innovation & Technology, Jupiter, FL, USA

Montina Van Meter

8. Department of Chemistry and Chemical Biology, TU Dortmund University, Dortmund, Germany

Herbert Waldmann

Contributions

M.D.D. conceived and directed the study. H.W. and S.S. provided the compound-screening platform to identify leads with H.S.H. and M.G.C.; F.G., T. Wegner and T.O.P. provided compounds (imidazolium salts) for the screening platform and additional derivatives for study in SAR. M.G. performed AbsorbArray and 2DCS studies. Y.T., E.L., Y.L. and X.L. performed analysis of the distribution of secondary structures in the RNA library, selected structures from 2DCS selections and those in Inforna. X.L. conducted synthesis and characterization of all miR-155-related compounds, in vitro cleavage gels, and cell-based assays in MDA-MB-231 and MCF-10a cells for miR-155. H.S.H. conducted annotation of the human miRNome, the chemoinformatic and bioinformatic analyses and all assays in HUVECs except for western blotting (performed by X.L.). R.I.B. conducted assays in CFPAC1 and MDA-MB-231 cells and validated qPCR primers in these cells. Y.L. validated the primers, performed RT–qPCR

analysis of miR-155-targeting compounds in MCF-10a cells, conducted synthesis and characterization of all MYC-related compounds and all cell-based assays for MYC unless otherwise noted. B.M.S. conducted synthesis and characterization of all JUN-related compounds, and all cell-based assays for JUN. J.L.C., C.Y. and W.L. conducted all cell-based experiments in Namalwa, HL-60 and Raji cells for MYC except for RT–qPCR. G.C. conducted bioinformatics analyses to generate LOGOs. D.A. and A.A. performed global proteomics analysis in MDA-MB-231 and HeLa cells. T. Wang and Q.M.R.G. assisted in the synthesis and characterization of compounds used in this study. Y.T. performed all in vitro binding assays for miR-155, in vitro fluorescence-based cleavage assays, all miRNA and RNA-seq analysis and in vivo studies. M.V.M. performed H&E staining of mouse lung tissue. J.L.C.-D. wrote the manuscript with assistance from all of the authors.

Corresponding authors

Correspondence to Frank Glorius, Herbert Waldmann or Matthew D. Disney.

Ethics declarations

Competing interests

M.D.D. is a founder of and consultant for Expansion Therapeutics. M.D.D. and J.L.C.-D. are founders of Ribonaut Therapeutics. Various aspects of these studies are the subject of a patent including US20160188791/WO2015021415 and US20220073910/WO2020167811 as well as disclosure numbers S-T00010US003 and S-T00014US002.

Peer review

Peer review information

Nature thanks Alessio Ciulli, Xiang (Shawn) Zhang and the other, anonymous, reviewer(s) for their contribution to the peer review of this

work.

Additional information

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Extended data figures and tables

Extended Data Fig. 1 To-Pro-1 dye displacement assay validation and newly identified small molecules that bind RNA are chemically dissimilar from previously known RNA binders.

a, To-Pro assay optimization and validation. Optimization of the signal to noise ratio of To-Pro-1 dye displacement from an RNA 3×3 internal loop library (3×3 ILL) by altering the concentration of the dye and the RNA (left; $n = 3$ independent replicates). Validation using Hoechst 33258 as a positive control to displace To-Pro-1 (middle; $n = 3$ independent replicates). Z-factor analysis of the screening conditions using 10 and 100 µM Hoechst 33258 (right). For each concentration, $n = 690$ in a single independent experiment. Data are reported as the mean ± SD. **b**, AbsorbArray screening with a 5'-^{32}P-labelled 3×3 ILL. Compounds were delivered to the array surface in dose response (200 nL of 10 – 0.625 mM, log_2 dilutions). **c**, Heat map of the Tanimoto analysis of 404 molecules found within Inforna show that **C1–C20** are chemically dissimilar to known RNA-binding matter. **d**, Average chemical similarity of **C1**–**C20** to compounds within the R-BIND database ($n = 104$ unique compounds). **e**, Comparison of the physicochemical properties of **C1–C20** and molecules housed within Inforna. The compounds identified herein are chemically distinct, having ~35-fold higher CLogP, an ~20% reduction in polar surface area (TPSA), and an ~15% reduction in the number of both H-bond donors and acceptors[15,57,58]. **f**, Unique chemical patterns indicate chemically similar compounds within the azolium scaffold. Compounds **C7–C10** have a Tanimoto scores ranging from 0.7 to >0.995 and cluster together. Compounds **C12**–**C15** are nearly chemically identical as they share the

cholesterol-derived azolium core, differing only by the other N-substituent. However, when compared to the remaining compounds (**C1–C11** and **C16–C20**), their Tanimoto score ranges from 0.29–0.32, indicating they are unique among the hits obtained. **C1** and **C11** exhibit high similarity (Tanimoto coefficient = 0.82), although they appear very different structurally. This could be due to their similar spatial orientations, as both have alkyl substituted benzenes on their azolium cores. Compounds **C4**, **C5**, **C6**, and **C20** are structurally unique compared to all other hits.

Source data

Extended Data Fig. 2 LOGO analyses for C1 – C6, showing both preferred and discriminated sequences and structures.

a, LOGO analysis of the selected RNAs in the top 0.5% of Z_{obs} and the secondary structures of the selected RNAs with the four highest Z_{obs} values. **b**, LOGO analysis of RNAs discriminated against by each small molecule, *i.e.*, in the bottom 0.5% of Z_{obs}, and the secondary structures of the RNAs discriminated against with the four lowest Z_{obs} values. Interestingly, many have GC closing base pairs and are 2×2 nucleotide internal loops. Similarities and differences were found amongst the small molecules. For example, within the azolium scaffold, **C1** prefers C at positions one and three and A at positions four and six (that is, internal loops with two Cs opposite two As), while **C3**, which differs by the substituents on the azolium core, prefers A at positions one, three and four, but C at positions six. Interestingly, **C1**–**C6**, despite their chemical differences in scaffolds, all discriminate against a 5’ GC closing base pair (G and C at positions one and six, respectively).

Extended Data Fig. 3 In vitro characterization of pre-miR-155-binder and its derivatives for pre-miR-155’s A bulge.

a, Representative binding curves of **pre-miR-155-binder** and two A bulges identified from the 2DCS selection and a fully based paired control, as determined by microscale thermophoresis (MST). The bulge found in pre-miR-155 is highlighted in green. **b**, Representative binding curves for **pre-**

miR-155-binder and its derivatives for a minimized A bulge construct or fully paired mutant, as determined by MST. **c**, Representative binding curves for **pre-miR-155-binder** and its derivatives and pre-miR-155's A bulge labelled with 2-aminopurine (2AP) or a fully base paired control (n = 2 independent experiments). **d**, Summary of K_ds of **pre-miR-155-binder** analogues and other imidazolium salts identified by 2DCS. The affinity of **C19** for the A bulge was undetermined due to its aggregation under assay conditions. **e**, Wild type pre-miR-155 but not the AU based paired mutant RNA can compete the binding of **pre-miR-155-RiboTAC** (5 µM) to 2AP labelled RNA (1 µM) with a K_d of 2.2 ± 0.9 µM (n = 2 independent experiments). **f**, **pre-miR-155-RiboTAC** induces in vitro RNA cleavage with wild-type (WT) pre-miR-155 construct (left), while no cleavage with mutated base pair control (right) (n = 3 independent experiments). All data are reported as the mean ± S.D. of independent replicates. All p-values were calculated using a two-tailed Student's t-test.

Source data

Extended Data Fig. 4 LOGO analyses for preferred and discriminated sequences for C7–C13.

Interestingly, ten of the 14 azoliums share a tetradecahydro-1*H*-cyclopenta[a]phenanthrene substituent, six of which are fused to the azolium (Supplementary Table 1). In general, all 7 compounds prefer loops with Cs and As, with a strong preference for A in the sixth position, with degree of preference at each position unique to each compound. Likewise, all 7 compounds discriminate against loops with Us, particularly in the sixth position. **Left**, LOGO analysis of the selected RNAs in the top 0.5% of Z_{obs} and the secondary structures of the selected RNAs with the four highest Z_{obs} values for compounds **C7**–**C13**. Generally, these compounds prefer motifs rich in C and A residues, with position 6 almost exclusively A. Similar to **C1**–**C6**, 3×3 internal loops predominate for **C7**–**C20** (75%) ($p < 0.001$, as compared to its distribution in 3×3 ILL), with single nucleotide bulges and 1×1 and 2×2 internal loops comprising 14%. Interestingly, **C7** and **C11** prefer to bind AU base pairs (pink box). **Right**, LOGO analysis of RNAs discriminated against by each small molecule, *i.e.*, in the bottom

0.5% of Z_{obs}, and the secondary structures of the RNAs discriminated against with the four lowest Z_{obs} values for compounds **C7**–**C13**. Motifs that are discriminated against for binding are significantly enriched for single nucleotide bulges (18%) ($p < 0.001$, as compared to its distribution in 3×3 ILL) and 1×1 nucleotide internal loops (29%) ($p < 0.001$, as compared to its distribution in 3×3 ILL), with position 6 predominately U. Notably, **C1** lacks the strong preference for A at position 6.

Extended Data Fig. 5 LOGO analyses for preferred and discriminated sequences for C14 – C20.

Left, LOGO analysis of the selected RNAs in the top 0.5% of Z_{obs} and the secondary structures of the selected RNAs with the four highest Z_{obs} values for compounds **C14**–**C20**. Generally, these compounds prefer motifs rich in C and A residues, with position 6 almost exclusively A. Similar to **C1**–**C6**, 3×3 internal loops predominate for **C7**–**C20** (75%) ($p < 0.001$, as compared to its distribution in 3×3 ILL), with single nucleotide bulges and 1×1 and 2×2 internal loops comprising 14%. **Right**, LOGO analysis of RNAs discriminated against by each small molecule, *i.e.*, in the bottom 0.5% of Z_{obs}, and the secondary structures of the RNAs discriminated against with the four lowest Z_{obs} values for compounds **C1–C20**. Preferred sequences for **C14–C20** are rich in C and A, with positions 1 and 6 preferred as G and A, respectively. Two of the azoliums, **C16** and **C17**, are the most similar to **C1**; in **C16**, the substituted phenyls in **C1** are replaced with methyl groups, while in **C17**, the alkyl chains have been shortened from C_{15} to C_7 and the 1,3-diisopropyl-2-methyphenyl groups have been replaced with benzyl groups. Like the other azolium salts, LOGO analyses of these three compounds revealed a preference for C and A nucleotides, with positional differences for all three molecules.

Extended Data Fig. 6 pre-miR-155-binder and derivatives directly engage pre-miR-155 and other A bulge containing miRNA precursors in MDA-MB-231 cells and in vitro.

a, The Chem-CLIP probe for target validation studies (**pre-miR-155-Chem-CLIP**) was synthesized by conjugating **pre-miR-155-COOH** to chlorambucil (CA; cross-linking module; blue triangle) and biotin (pull-down module; yellow circle). A control Chem-CLIP probe, Ac-CA-biotin, comprises the cross-linking and pull-down module but lacks the RNA-binding module and accounts for non-selective reaction of the CA module. **b**, Chem-CLIP is a proximity-based reaction driven by the RNA-binding module. Following treatment, total RNA is isolated and purified with streptavidin beads and analysed by RT-qPCR. **c**, Results of target validation studies in MDA-MB-231 breast cancer cells. Cells were treated with either Ac-CA-Biotin (100 nM) or **pre-miR-155-Chem-CLIP** (100 nM). Enrichment of pre-miR-155 in the pulled-down fraction, as compared to the starting cell lysate, was quantified by RT-qPCR (left). Co-treatment with **pre-miR-155-Chem-CLIP** and **pre-miR-155-binder** resulted in dose dependent depletion of pre-miR-155 in the pull-down fraction, indicating that the Chem-CLIP probe and the simple binding compound **pre-miR-155-binder** occupy the same site ($n = 3$ biological replicates). **d**, **pre-miR-155-Chem-CLIP** cross-links to pre-miR-155 at a nucleotide 5 bp from the A bulge binding site (indicated with a blue arrow in the secondary structure), identified by in vitro Chem-CLIP-Map-Seq. **e**, Effect of **pre-miR-155-binder** (red; left) and **pre-miR-155-amide-binder** (blue; middle) ($n = 3$ biological replicates) on pri-, pre-, and mature miR-155 levels as well as the two compounds' effects on cellular viability (right) ($n = 4$ biological replicates). **f**, Chem-CLIP analysis (100 nM **pre-miR-155-Chem-CLIP**) of miRNA precursors that have the same bulge as pre-miR-155 bound by **pre-miR-155-binder** ($n = 3$ biological replicates) where ND indicates that the miRNA was not detectable in the pulled down fraction. Effect of **pre-miR-155-amide-binder** (0.1 to 10 μM) on the levels of miRNAs that contain the A bulge in miR-155, as determined by RT-qPCR ($n = 3$ biological replicates). **Pre-miR-155-RiboTAC** (100 nM) has no effect on the levels of mature RNAs that share **C1**'s A bulge in miR-155 (Extended Data Fig. 6h). All data are reported as the mean ± S.D. All p-values were calculated using a two-tailed Student's t-test.

Source data

Extended Data Fig. 7 In vitro and cellular activities of pre-miR-155-RiboTAC and its derivatives.

a, **left**: Representative gel autoradiogram of the cleavage of 5′-[^{32}P]-pre-miR-155 by RNase L induced by **pre-miR-155-RiboTAC** and quantification thereof (IC_{50} = ~100 nM; *n* = 3 independent replicates); **right**: Representative gel autoradiogram of the **pre-miR-155-RiboTAC**-induced cleavage of pre-miR-155 as a function of **pre-miR-155-binder** concentration (IC_{50} = 1.9 ± 0.5 μM), demonstrating **pre-miR-155-binder** and **pre-miR-155-RiboTAC** bind the same site, and the corresponding quantification thereof (*n* = 3 independent replicates). **b**, Representative gel autoradiogram of (left to right): RNase L cleavage of cleavage site mutant of pre-miR-155 induced by **pre-miR-155-RiboTAC**; RNase L cleavage of the binding site mutant pre-miR-155 induced by **pre-miR-155-RiboTAC**; and RNase L cleavage of WT pre-miR-155 induced by **Ac-RiboTAC**, which lacks the RNA-binding module (*n* = 3 independent replicates). For panels **a** and **b**, "H" indicates a hydrolysis ladder while "T1" indicates treatment with RNase T1, which cleaves guanosine nucleotides. **c**, Effect of **pre-miR-155-RiboTAC** (*n* = 3 biological replicates) or **pre-miR-155-Ctr** (*n* = 4 biological replicates) on the viability of MDA-MB-231 cells following a 48 h treatment, as assessed by WST-1 assay. **d**, **pre-miR-155-RiboTAC** (single dose) reduces mature miR-155 levels in MDA-MB-231 cells in a time-dependent manner (*n* = 3 biological replicates). **e**, Duration of **pre-miR-155-RiboTAC**'s effect on pre-miR-155's levels after a 48 h treatment (wash out experiment; *n* = 3 biological replicates). The compound-containing medium was removed after the treatment period, and total RNA was harvested at the indicated times post-treatment, affording a $t_{1/2}$ = 19.3 h. The observed half-life is in accord with reported half-lives of miR-155, which range from 12.4–28 h[15]. **f**, Effect of **Ac-RiboTAC** (green), the RNase L recruiter lacking the RNA-binding module, and **pre-miR-155-Ctr** (red), a chimera with the less active regioisomer of the RNase L recruiter, on miR-155 biogenesis after a 48 h treatment as determined by RT-qPCR (*n* = 3 biological replicates). **g**, Effect of **pre-miR-155-RiboTAC** and **pre-miR-155-binder** on pre-miR-155 (left; *n* = 6 biological replicates) and mature miR-155 (right; *n* = 3 biological replicates for compound-treated cells; *n* = 6 biological replicates for vehicle-treated) levels in CFPAC-1

(pancreatic ductal adenocarcinoma) cells, as determined by RT-qPCR. **h**, Correlation of miRNome-wide changes of MDA-MB-231 cells treated with **pre-miR-155-RiboTAC** (100 nM) or **LNA-155** (50 nM) (*n* = 3 biological replicates). Comparison of normalized read counts for each miRNA between **pre-miR-155-RiboTAC** and **LNA-155** treatment (left). Effect of **pre-miR-155-RiboTAC** (100 nM, *n* = 3 biological replicates) on other miRNAs sharing the targeted A bulge (right). **i**, Correlation of transcriptome-wide changes of MDA-MB-231 cells treated with **pre-miR-155-RiboTAC** (100 nM) or **LNA-155** (50 nM) (*n* = 3 biological replicates). From left to right: (i) Volcano plot of transcriptome-wide changes of MDA-MB-231 cells treated with **pre-miR-155-RiboTAC** (100 nM) *vs*. vehicle and **LNA-155** (50 nM) *vs*. vehicle after a 48 h treatment period; (ii) comparison of normalized read counts for each gene between **pre-miR-155-RiboTAC** and **LNA-155** treatment; (iii) comparison of normalized % change for genes commonly affected by both **pre-miR-155-RiboTAC** and **LNA-155** treatment (*n* = 26 genes, right). All data are reported as the mean ± S.D. All p-values were calculated using a two-tailed Student's t-test.

Source data

Extended Data Fig. 8 Cellular and *in vivo* activity of pre-miR-155-RiboTAC and its derivatives.

a, Effect of **pre-miR-155-RiboTAC** (100 nM) on the miRNome of MDA-MB-231 cells forced to express WT pre-miR-155 (left; *n* = 3 biological replicates); and effect of **pre-miR-155-RiboTAC** (100 nM) on the miRNome of MDA-MB-231 cells forced to express the binding site mutant of pre-miR-155 (right; *n* = 3 biological replicates). **b**, Effect of **pre-miR-155-Ctr** on the migration of MDA-MB-231 cells (*n* = 3 biological replicates). **c**, Drug Metabolism and Pharmacokinetics (DMPK) analysis of **pre-miR-155-RiboTAC** in C57BL/6 mice (*n* = 3 mice per time point). **d**, In vivo treatment of **pre-miR-155-RiboTAC** (1 mg/kg, *q.o.d.*, 30 days) decreased the number of lung nodules (nodules indicated in red) stained with Bouin's solution (*n* = 5 mice). **e**, Lung tissue treated with **pre-miR-155-RiboTAC**, but not **pre-miR-155-amide binder**, exhibited decreased mature miR-155 levels, as determined by FISH (*n* = 5 mice). All data are

reported as the mean ± S.D. Statistical significance was indicated by a Walds test (**a**) or two-tailed Student's t-test (**b-e**).

Source data

Extended Data Fig. 9 Cellular activity of pre-miR-155-RiboTAC in MCF-10a cells overexpressing miR-155 and in HUVECs.

a, Effect of **pre-miR-155-RiboTAC** on pre- and mature miR-155 in MCF-10a cells forced to express wild-type pre-miR-155, a binding site mutant of pre-miR-155 (A bulge to AU base pair), or a RNase L cleavage site mutant (pyrimidine-rich asymmetric loop mutated to base pairs) (*n* = 3 biological replicates). **b**, Forced expression of pre-miR-155, a mutant pre-miR-155 in which **pre-miR-155-binder**'s binding site has been mutated, or a mutant pre-miR-155 in which the RNase L cleavage site has been abolished enhances the migratory nature of MCF-10a cells, a model of healthy breast epithelium (*n* = 3 biological replicates). **c**, **pre-miR-155-RiboTAC** impairs the migration of MCF-10a cells forced to express WT pre-miR-155 but not those forced express the binding site or cleavage site mutants (*n* = 3 biological replicates). **d**, **left**: Effect of **pre-miR-155-RiboTAC** on mature miR-155 levels in HUVECs, as determined by RT-qPCR (*n* = 3 biological replicates). **right**: Effect of **pre-miR-155-binder** on the cleaving capacity of **pre-miR-155-RiboTAC**, as assessed by measuring pre-miR-155 levels by RT-qPCR (*n* = 4 biological replicates). **e**, Effect of **pre-miR-155-binder** on mature and pre-miR-155 levels in HUVECs, as determined by RT-qPCR (*n* = 3 biological replicates for mature miR-155; *n* = 4 biological replicates for pre-miR-155). **f**, Effect of control RiboTAC **pre-miR-155-Ctr** on mature and pre-miR-155 levels in HUVECs, as determined by RT-qPCR (*n* = 3 biological replicates). **g**, Effect of **pre-miR-155-RiboTAC** (100 nM) on the levels of mature miRNAs that harbour miR-155's A bulge in their precursors in HUVECs, as determined by RT-qPCR (*n* = 3 biological replicates). **h**, Representative Western blot and quantification thereof of the effect of **pre-miR-155-RiboTAC** on pre-miR-155's downstream target, Von-Hippel Lindau (VHL) (*n* = 3 biological replicates). **i**, Effect of **pre-miR-155-RiboTAC** on the angiogenic capacity of HUVECs (*n* = 2

biological replicates). All data are reported as the mean ± S.D. All p-values were calculated using a two-tailed Student's t-test.

Source data

Extended Data Fig. 10 In vitro and in cellular characterization of c-Jun-binder and its derivatives.

a, Representative binding curves to measure the affinity of **c-Jun-binder**, **c-Jun-RiboTAC**, and **c-Jun-Ctr** for a model of the c-*Jun* RNA where the A bulge has been replaced with the fluorescent adenine mimic, 2-aminopurine (2AP) (*n* = 2 independent experiments). **b**, Representative binding curves to measure the affinity of **c-Jun-binder**, **c-Jun-RiboTAC** and **c-Jun-Ctr** for a Cy5-labelled model of the c-*Jun* IRES and a fully paired RNA control (*n* = 2 independent experiments). The fully paired RNA was created by mutating the A bulge binding site to an AU pair. **c**, Chemical structure of **c-Jun-Chem-CLIP** probe and control probe. Pull-down of the c-*Jun* mRNA in vitro by **c-Jun-Chem-CLIP** or **Ctr-Chem-CLIP** (*n* = 3 replicates). **d**, **left**: Target engagement studies completed by Chem-CLIP in MIA PaCa-2 cells (*n* = 3 biological replicates). Quantification of the enrichment of c-*Jun* mRNA by **c-Jun-Chem-CLIP** or **Ctr-Chem-CLIP** in pulled-down fraction as compared to RNA not subjected to pull-down; **right**: Competitive Chem-CLIP experiments in which MIA PaCa-2 cells were co-treated with **c-Jun-Chem-CLIP** at a constant concentration and increasing concentrations of **c-Jun-binder**, which depletes c-*Jun* mRNA in the pull-down fractions in a dose dependent fashion (*n* = 4 biological replicates). **e**, Effect of **c-Jun-binder** on c-JUN protein levels in Mia-PaCa-2 cells (*n* = 3 biological replicates). **f**, **c-Jun-RiboTAC** induces in vitro RNA cleavage with WT c-*Jun* construct (left), while no cleavage of the mutated base pair control (right) was observed (*n* = 3). All data are reported as the mean ± S.D. All p-values were calculated using a two-tailed Student's t-test.

Source data

Extended Data Fig. 11 In vitro and cellular characterization of c-Myc-binder and its derivatives.

a, Representative binding curves to measure the affinity of **c-Myc-binder**, **c-Myc-RiboTAC** and **c-Myc-Ctr** for a Cy5-labelled model of the c-Myc IRES 5'U<u>UC</u>G/3'A<u>CC</u>C 2×2 internal loop or a fully paired RNA control (*n* = 2 independent experiments). **b**, Chemical structures of **c-Myc-Chem-CLIP** and **Ctr-Chem-CLIP** probes (top) and target validation studies in HeLa cells (bottom). **c-Myc-Chem-CLIP** dose dependently enriches *MYC* mRNA (lower left; *n* = 3 biological replicates), which is competed by **c-Myc-binder** (lower right; *n* = 6 biological replicates). **c**, Effect of **c-Myc-binder** on *MYC* mRNA and protein levels in HeLa cells (*n* = 3 biological replicates). **d**, **c-Myc-RiboTAC** induces in vitro cleavage of the WT *MYC* IRES (left), while no cleavage of a mutated base pair control was observed (middle). **Pre-miR-155-RiboTAC** and **c-Myc-RiboTAC** have no effect on the non-cognate RNA (right) (*n* = 3 biological replicates for all panels). Concentrations were selected based on the cellular activity for the cognate target. **e**, Chemical structures of **c-Myc-amide-binder** and **Ac-RiboTAC**. **f**, Effect of **c-Myc-amide-binder**, **Ac-RiboTAC**, and **c-Myc-Ctr** on *MYC* mRNA and protein levels in HeLa cells (*n* = 3 biological replicates). **g**, Effect of siRNA knock-down of RNase L (~85%) on the activity of **c-Myc-RiboTAC** in HeLa cells (*n* = 3 biological replicates). **h**, Effects of **c-Myc-RiboTAC** and BRD4 degrader **MZ1** on the proliferation and apoptosis of HeLa cells upon 48 h treatment (*n* = 3 biological replicates). **i**, Effect of **c-Myc-RiboTAC** on *MYC* mRNA levels (left; n = 3 biological replicates), protein abundance (middle; (n = 2 biological replicates for vehicle (note that the third lane was excluded from quantification due to air bubble) and n = 3 biological replicates for compound-treated sampled), and proliferation (right; *n* = 3 biological replicates) in MDA-MB-231 cells. **j**, Effect of siRNA knockdown of RNase L on **c-Myc-RiboTAC**-mediated degradation of *MYC* mRNA in MDA-MB-231 cells (*n* = 3 biological replicates). **k**, Effect of **c-Myc-RiboTAC** on a *c-Myc*-IRES-luciferase reporter or a binding site mutant thereof in transfected HEK293T cells (*n* = 4 biological replicates). **l**, Transcriptome-wide changes observed in HeLa cells treated with vehicle, **c-Myc-RiboTAC**, scrambled siRNA, or *MYC* siRNA (*n* = 3 biological replicates). Volcano plots of transcriptome-wide changes of HeLa cells treated with *MYC*-siRNA (1 nM) *vs*. scrambled siRNA (1 nM) after a 48 h treatment period (left). Cumulative curves of *MYC* targets and HIF-1α targets in HeLa cells treated with **c-Myc-RiboTAC** (10 μM) *vs*. vehicle (middle) or c-Myc-siRNA (1 nM) *vs*. scrambled siRNA (1 nM) (right). **m**,

Volcano plots of proteome-wide changes of HeLa cells treated with **c-Myc-RiboTAC** (10 µM) *vs*. vehicle (left) or *MYC*-siRNA (1 nM) *vs*. scrambled siRNA (1 nM) (right) after a 48 h treatment period (*n* = 3 biological replicates). **n**, Abundance of RNase L mRNA in various cell lines (left; *n* = 3 biological replicates). Effect of **c-Myc-Ctr** (10 µM) and **c-Myc-RiboTAC** (10 µM) on *MYC* mRNA levels in lymphoma or leukaemia cells, Raji, and HL-60 cells, respectively (middle left; *n* = 6 biological replicates for vehicle and *n* = 3 biological replicates for compound treatment) and on apoptosis (middle right and right; *n* = 2 biological replicates). Interestingly, these data suggest that RiboTAC activity is correlated with RNase L expression. All data are reported as the mean ± S.D. of biologically independent replicates. Statistical significance was calculated using a two-tailed Student's t-test (**b-h, n**), Wald's test (**i, left and m**), or Kolmogorov–Smirnov test (**i, middle and right**).

Source data

Supplementary information

Supplementary Information

Supplementary Tables 1 and 2; Supplementary Figures 1–9; Supplementary Methods; and synthetic methods and characterizations.

Reporting Summary

Source data

Source Data Fig. 2

Source Data Fig. 3

Source Data Fig. 4

Source Data Fig. 5

Source Data Extended Data Fig. 1

Source Data Extended Data Fig. 3

Source Data Extended Data Fig. 6

Source Data Extended Data Fig. 7

Source Data Extended Data Fig. 8

Source Data Extended Data Fig. 9

Source Data Extended Data Fig. 10

Source Data Extended Data Fig. 11

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Reprints and Permissions

About this article

Cite this article

Tong, Y., Lee, Y., Liu, X. *et al.* Programming inactive RNA-binding small molecules into bioactive degraders. *Nature* **618**, 169–179 (2023). https://doi.org/10.1038/s41586-023-06091-8

- Received: 11 August 2021
- Accepted: 17 April 2023
- Published: 24 May 2023
- Issue Date: 01 June 2023
- DOI: https://doi.org/10.1038/s41586-023-06091-8

Share this article

Anyone you share the following link with will be able to read this content:

Get shareable link

Sorry, a shareable link is not currently available for this article.

Copy to clipboard

Provided by the Springer Nature SharedIt content-sharing initiative

---

This article was downloaded by **calibre** from https://www.nature.com/articles/s41586-023-06091-8